University of Bath

PHD

Simulation of semiconductor devices

Oriato, Diego

Award date:
2001

Awarding institution:
University of Bath

Link to publication

Alternative formats
If you require this document in an alternative format, please contact:
openaccess@bath.ac.uk

Copyright of this thesis rests with the author. Access is subject to the above licence, if given. If no licence is specified above, original content in this thesis is licensed under the terms of the Creative Commons Attribution-NonCommercial 4.0 International (CC BY-NC-ND 4.0) Licence (https://creativecommons.org/licenses/by-nc-nd/4.0/). Any third-party copyright material present remains the property of its respective owner(s) and is licensed under its existing terms.

Take down policy
If you consider content within Bath's Research Portal to be in breach of UK law, please contact: openaccess@bath.ac.uk with the details. Your claim will be investigated and, where appropriate, the item will be removed from public view as soon as possible.

Download date: 05. May. 2023

Simulation of Semiconductor Devices

Diego Oriato

September 2001

Submitted in partial fulfilment of the requirements for the degree of
Doctor of Philosophy at the University of Bath, Bath, England

©This copy of the thesis has been supplied on condition that anyone who consults it is understood to recognise that its copyright rests with the author and that no quotation from the thesis, nor any information derived therefrom, may be published without the authors prior written consent.

UMI Number: U145918

All rights reserved

INFORMATION TO ALL USERS
The quality of this reproduction is dependent upon the quality of the copy submitted.

In the unlikely event that the author did not send a complete manuscript
and there are missing pages, these will be noted. Also, if material had to be removed,
a note will indicate the deletion.

UMI U145918
Published by ProQuest LLC 2013. Copyright in the Dissertation held by the Author.
Microform Edition © ProQuest LLC.
All rights reserved. This work is protected against
unauthorized copying under Title 17, United States Code.

ProQuest LLC
789 East Eisenhower Parkway
P.O. Box 1346
Ann Arbor, MI 48106-1346

UNIVERSITY OF BATH
LIBRARY
27 FEB 2002
Ph.D

Abstract

In this thesis a drift diffusion model coupled with self-consistent solutions of Poisson's and Schrödinger's equations, is developed and used to investigate the operation of Gunn diodes and GaN-based LEDs. The model also includes parameters derived from Monte Carlo calculations of the simulated devices. In this way the characteristics of a Monte Carlo approach and of a quantum solver are built into a fast and flexible drift-diffusion model that can be used for testing a large number of heterostructure designs in a time-effective way. The full model and its numerical implementation are described in chapter 2.

In chapter 3 the theory of Gunn diodes is presented. A basic model of the dynamics of domain formation and domain transport is described with particular regard to accumulation and dipole domains. Several modes of operation of the Gunn device are described, varying from the resonance mode to the quenched mode. Details about transferred electron devices and negative differential resistance in semiconductor materials are given.

In chapter 4 results from the simulation of a simple conventional gunn device confirm the importance of the doping condition at the cathode. Accumulation or dipole domains are achieved respectively with high and low doping densities. The limits of a conventional Gunn diode are explained and solved by introducing the heterostructure Gunn diode. This new design consists of a conventional GaAs transit region coupled with an electron launcher at the cathode, made using an AlGaAs heterostructure step. Simulations show the importance of the insertion of a thin highly-doped layer between the transit region and the electron launcher in order to improve device operation.

Chapter 5 is an introduction to III-nitrides, in particular GaN and its alloy InGaN. We outline the discrepancy in the elastic and piezoelectric parameters found in the literature. Strain, dislocations and piezoelectricity are presented as the main features of a InGaN/GaN system.

In chapter 6 an extensive simulation of the dependence of the optical band gap of a single InGaN quantum well on the piezoelectric and spontaneous polarization is

reported. Quantum Confined Stark Effect and screening mechanisms are found to play a major role. The simulation of a novel InGaN/GaN double quantum well LED is presented. A wide well is used to capture electrons that tunnels in a narrow well where they recombine with holes. Resonant asymmetric tunneling of electron and holes is used to increase the efficiency of the device. Our simulations show that piezoelectric field deeply affects the performance of the device and we outline the optimal device conditions.

To my family

Mamma, Papa', Cristina, Elena, Massimiliano and my dearest nephew Leonardo.

Acknowledgements

Firstly I would like to thank Alison Jensen for giving me the opportunity of being part of her research group, providing vast resources, unlimited help and precious support in every occasion. I would also like to thank professor Wang at the University of Bath for his financial help, experimental data and useful discussions.

A big hug and unlimited thanks go to all the present members of the modeling group at the University of Bath. This especially includes Simon Martin that has been of invaluable help in many stimulating conversations on the nature of the matter as well as on the matter of the nature. Thanks to Simon also for relieving the frustration of long days spent debugging computer programs or fixing workstations. I would like to thank Agapi Kambili for making me feel in a more international environment and for having taught me precious rules of Physics and of Life. Thanks to Fulian Qiu for making me feel useful. Thanks to Michael Cass for sharing the office with me and bearing all my oddities. Thanks to Matthew Webster who always innocently believes I was right.

A special thanks goes to Christopher Blades who protected me like a child until he got one of his own. Without Chris I think I would have not been able to complete this work. He helped me from the very beginning and taught me the precious art of rebooting a unix machine. I would also like to thank him for being just a simple and true friend and a wonderful person.

There have been many incredible people who made the last three years something to remember. Among them, I would like to thank Eva Ashford for being incredibly kind and patient with me and Nikki Broomley for teaching me the art of swimming. Thanks to Amir Atiabi for being the most joyful person I have ever met. Thanks to Jean-Philippe Cross, Hadi Madani, Michael Gill-Carey and Arach Goldar, my unforgettable house mates. A very special thanks goes to Claudia Neri who gave me force and confidence and to whom I am forever debtor.

Last but by no means least I would like to thank my family for that great and sometimes not deserved supply of love that I have been always receiving from them despite the long distance that divides us.

Contents

Chapter 1

Introduction

1.1 Semiconductor Device Simulation

Modelling and simulation of modern semiconductor devices and circuits drastically reduces development costs and time to market of new technologies. As an example, in the 1995, the direct benefit of modelling to AT&T Corporation was estimated to be 10-15 million of US dollars per year, and each design was on average made available on the market 1.5 years earlier than it would have been with experimental development alone [1].

The expected benefits of simulation are:

- the ability to optimise existing designs and to characterise devices without building them,
- the ability to explore alternate and novel designs in a timely and flexible way,
- the ability to predict difficult-to-measure quantities.

In order to fulfil these requirements the primary use of semiconductor simulation is to understand what happens in semiconductor material and devices. The

models have to be based on true physical characteristics but there also has to be a trade off between efficiency in use and completeness of the physical description.

In the past the problem of carrier transport dynamics in semiconductors has been mainly solved using the Boltzmann transport equation, drift diffusion equations of electrons and holes, hydrodynamic models and energy balance equations [2]. These models are a classical description of the physics of semiconductors. Their solution involves complex non-linear equations in three dimensions that are quickly solved by modern computers.

Nowadays, the quest for fast and short devices and the improvement of the growth of heterostructures has changed drastically the scenario of the device semiconductor technology. Two important technologies of crystal growth have evolved from the vapour phase epitaxy introduced in 1960, molecular beam epitaxy (MBE) and metal organic chemical vapour deposition (MOCVD), which make it possible to grow lattice-matched semiconductor structures having a characteristic dimension of few nanometres and less. Nano-structures and quantum electronics effects are becoming important and the previous classical models are no longer sufficient. Monte Carlo approaches [3], Schrödinger equation solvers and quantum transport models [4] have been developed in order to describe new effects in the new devices. But with the increase in the complexity of these new models, we are witnessing an exponential growth in the computing time and a strong reduction in the flexibility of these methods.

In this work, we present a versatile one-dimensional time-dependent numerical technique for modelling any kind of semiconductor structure, with particular regard to Gunn diodes or transferred electron devices (TEDs) whose theory is discussed in detail in chapter 3, and GaN-based LEDs as described in chapter 5. Our technique involves a self-consistent solution of Poisson and Schrödinger equation and carrier continuity equations, known as Drift-Diffusion Model (DDM). In contrast with a Monte Carlo approach, the DDM allows a reduction in computer time and resources. This permits more rapid testing of different semiconductor structures. The DDM and its numerical implementation are reported in chapter 2. In order to simulate quantum well structures and tunneling barriers, the one dimension time independent Schrödinger equation is solved. Wavefunctions and eigenvalues are obtained by respectively the Transfer Matrix Technique (TMT) and the Transmission Coefficient approach. These techniques will be described

in details in chapter 2. Chapter 2 also describes the numerical issues encountered in the computation stage. We describe several settings that have showed better performance as far as the speed and stability of the solvers are concerned. Different optimal conditions were obtained for different types of devices. As an example, in the GaN-LED simulation, we had to deal with numerical issues related to the wide band gap nature of GaN. We must remember that modeling is a multi-disciplinary science and besides the usual physical problems, other issues more related to mathematics and computer science come into play. For the sake of brevity, we have omitted parts of the research that have been dedicated to the solution of numerical or computing problems even if, in terms of time scale, such complications have kept us occupied as much as the physics problems. We have excluded for obvious reasons the 10,000 lines computer code (in FORTRAN and in C++) that represents the main contribution and the only tangible proof of the effort done.

As previously stated, the model we have developed has been used mainly in the simulation of two semiconductor device: the Gunn diode and the GaN-based LED.

1.2 The Gunn diode

The Gunn diode is one of the most common high-frequency solid state oscillators at present with a domain of applications covering 10 to over 300 GHz. Millimeter and submillimeter-wave applications [5] such as in intelligent cruise control/collision avoidance car radar and in local multipoint distribution systems (LMDS), require high frequency, high power, wide band width, low output noise and wide temperature operation Gunn diodes. New designs have been proposed mainly based on complex heterostructure devices [6]. Simulation has become the only and compulsory tool in order to understand the physics underneath these new structures

Our computer model has been initially tested on a 3 μm long n-GaAs sample at temperature T_0=300 K with two different doping inhomogeneities at the cathode and a 2.75 μm layer doped at $5{\cdot}10^{15}$ cm^{-3}. In both cases the semiconductor has been biased in its negative differential mobility region by a voltage of 2.1 V. In

Device A (see figure 1.1) a doping notch of $7{\cdot}10^{15}$ cm^{-3} at the cathode is a source of carriers. An accumulation-layer can form and propagate.

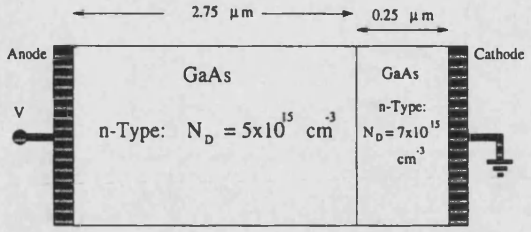

Figure 1.1: Device A:accumulation-layer mode

In Device B (see figure 1.2) the layer near the cathode has a lower doping level of $3.5{\cdot}10^{15}$ cm^{-3}. A depletion region forms and sets up a dipole region. These results are reported in chapter 4 and are in agreement with the literature on Gunn diodes [5] showing how the doping level in the notch determines the mode of operation.

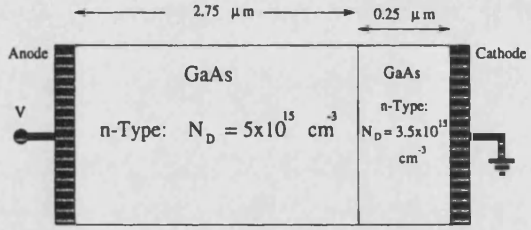

Figure 1.2: Device B: diode-layer mode

Particular attention has been given to the simulation of heterojunction Gunn diodes. For this purpose the drift-diffusion model has been enriched with a numerical method for analysing heterostructures [7] using a thermionic emission approach described in chapter 2. Chapter 4 presents the analysis and simulation of a graded-gap injector Gunn diode schematically drawn in figure 1.3.

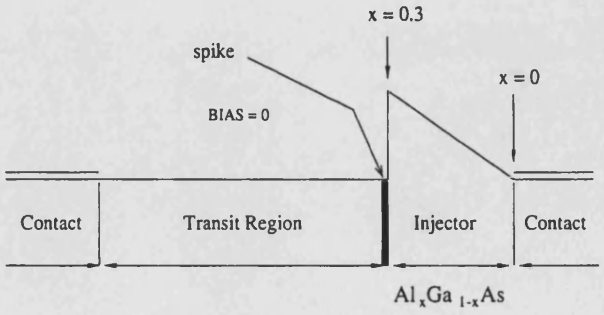

Figure 1.3: Graded-Gap Gunn Diode with n^+ spike

Self-sustained domain propagation under dc voltage has been obtained in good agreement with the experimentally measured natural frequency of oscillation of

similar devices [8]. For the same structure the insertion of a n^+ spike between the AlGaAs injector and the transit region has been found to be of great importance in terms of improving performance. It has been found that there is an optimum value for the doping of the n^+ spike; our simulation shows that this depends critically on the applied bias.

1.3 GaN-based LED

More than a quarter century has elapsed since the first demonstration of stimulated emission in GaN under optical pumping [9]. Progress has been slow because of intrinsic difficulties linked to the growth of GaN and the lack of appropriate substrates. For almost two decades, p-doping of GaN was not feasible. With the achievement of p-doping by an annealing process [11][12], devices based on GaN were demonstrated. Since the technological breakthroughs made by Nichia, such as a blue laser diode working at room temperature [13], in cw mode and lasting 10,000 hours, gallium nitride has been the subject of intense research activity world-wide. Devices based on GaN would have not been possible without the emergence of new approaches to the heteroepitaxy of GaN leading to layers exhibiting lower dislocation densities than those grown using conventional heteropitaxy.

However, even after more than 30 years of research, III-nitrides remain an *immature* semiconductor. The knowledge of many parameters for GaN and other nitrides is far from satisfactory, or at least less than that of other III-V semiconductors[14]. A major consequence of this is the impossibility of separating optical or electrical transport issues from the growth problems. Device simulation has to be able to include new physical phenomena and has to be flexible enough to include effects that are material or even sample dependent. In chapter 2 the theory of strain, piezoelectric polarization and traps is described and applied to our model.

We have initially simulated a InGaN/GaN single quantum well structure. Under pseudomorphic growth of the InGaN quantum well, we report in chapter 6 the dependence of the energy separation of the electron and hole ground levels in the quantum well (optical band gap) as a function of piezoelectric polarization. We

interpret our results within the Quantum Confined Stark Effect (QCSE) model even if screening mechanisms of the piezoelectric field have to be considered in order to obtain agreement with experimental results. The magnitude of the screening is shown to vary sensibly with the well width, the indium content in the quantum well and with the bias. Moreover, results reported in the literature presents different values of screening magnitude. A closer comparison of quantum wells with different widths is done. We study the problem of indium fluctuations and misfit dislocations in a quantum well and their effect on the optical band gap.

We then have simulated a novel double quantum well InGaN/GaN design, shown schematically in figure 1.4, that has already produced bright blue LEDs[15][16]. In this design, a wide well (WW) and an active narrow quantum well (NW) are coupled via a narrow barrier which is transparent for electrons and blocking for holes. The high cross-section WW can more effectively capture electrons injected from the left. Then electrons tunnel into the NW and stay there until they combine radiatively with holes. Holes, coming from the right contact, have a sufficiently heavy mass that they can be captured even by a small cross-section NW. A blocking barrier towards the WW then confines holes in the NW.

By adjusting the alloy composition in the WW and NW and choosing an appropriate NW width, it is possible to match the energy position of the bottom of the WW with the energy position of the sub-band minimum of the NW, thus allowing for the possibility of electron resonant tunnelling between the two wells. The heavy mass of the holes inhibits hole tunnelling from the NW to the WW and the energy separation between the hole fundamental sub-band in the NW and WW means that hole penetration into the WW is not possible without thermal activation. Thus, the main effect of the coupled wells is to increase the quantum efficiency for the LED.

Particular care has been taken to investigate the role of the piezoelectric field on the band profile of the device. Resonance conditions are found to be highly sensitive to small variations of piezoelectric field. Furthermore, build-up of strain-induced charge prevents optimal electron injection. For these reasons, different designs have been tested as shown in figure 1.4(b)(c) where a reduction of the piezoelectric field is obtained by grading the indium content in the WW. The device structure of figure 1.4(c) has shown the most promising characteristics.

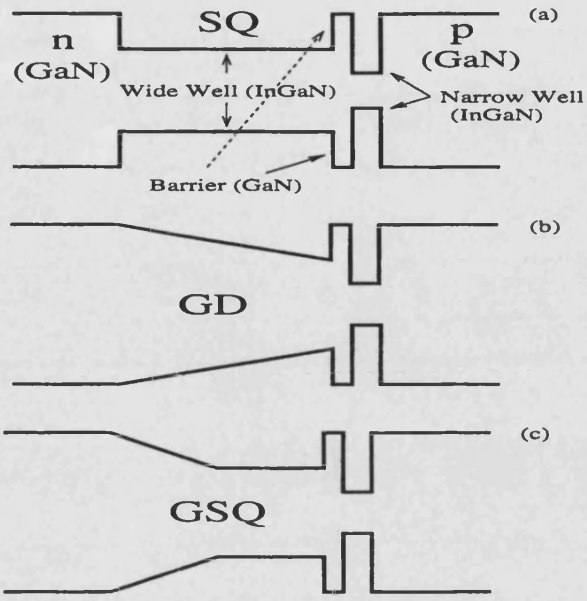

Figure 1.4: Schematic band profile for (a) the GaN-based CART LED with Square Wide Well, (b) the Graded Wide Well structure (GD), (c) the Graded and Square Wide Well structure (GSQ).

We have studied the effect of n-type doping in the WW and in the NW in order to screen the piezoelectric field. Increasing of overlap between wavefunctions has been achieved using high level of doping. Unfortunately, hole injection seems to become more effective due to a strong shift of the entire band profile of the device.

Lastly, we have investigated the role of electron traps on the IV characteristic of a simple GaN np-junction. This has been done in order to understand the impact of misfit dislocations and background doping of GaN on the carrier transport in III-nitrides. We found that space charge limited current saturation takes place for densities of traps above 10^{16} cm^{-3}.

References

[1] William M. Coughran, AT&T Bell Laboratories, *Modelling and Simulation: a NIST Multi-Laboratory Strategic Planning Workshop*, Gaithersburg MD, 1995.

[2] K. Hess, *Advanced theory of semiconductor devices*, Prentice-Hall (1988).

[3] C. Jacoboni and P. Lugli, *"The Monte Carlo Method for Semiconductor Device simulation"*, Springer-Verlag Wien, New York 1989, edited by Selberherr.

[4] C. Jacoboni, A. Bertoni, P. Bordone, R. Brunetti, *Mathematics and Computers in Simulation*, **55**, 67 (2001).

[5] G. B. Tait and C.M. Krowne, *Solid-State Electronics*, **30**, 1025 (1987).

[6] N. R. Couch, H. Spooner, P. H. Beton, M. J. Kelly, M. E. Lee, P. K. Rees and T. M. Kerr, *IEEE Electron Device Lett.*, **EDL-10**, 288 (1989).

[7] M. S. Lundstrom and R. J. Schuelke, *IEEE Transactions on Electron Devices*, **30**, 1151 (1983).

[8] H. Spooner and N.R. Couch, *GEC Journal of Research*, **7**, 34 (1989).

[9] R. Dingle, R. Shaklee, R. F. Leheny and R. B. Zetterstrom, *Appl. Phys. Lett.*, **19**, 5 (1971).

[10] K. Horio and H. Yanai, *IEEE Transactions on Electron Devices*, **37**, 1093 (1990).

[11] H. Amano, I. Akasaki, T. Kozawa, K. Hiramatsu, N. Sawaki, K. Ikeda and Y. Ishii, *J. Lumin.*, **40-48**, 121 (1988).

[12] H. Amano, M. Kito, K. Hiramatsu and I. Akasaki, *Jpn. J. Appl. Phys.*, **28**, L2112 (1989).

[13] S. Nakamura, G. Fasol, *The Blue Laser Diode: GaN based light emitters and lasers*, Springer (1997).

[14] S. C. Jain, *Appl. Phys. Rev.*, **87**, 965 (2000).

[15] Y. T. Rebane, Y. G. Shreter and W. N. Wang, *Appl. Surf. Science*, 6313 (2000).

[16] D. Oriato, A. B. Walker and W. N. Wang, *VLSI Design, Proceedings of IWCE Glasgow*, in press (2000).

Chapter 2

Method

2.1 Introduction

The problem of classic carrier transport dynamics in the semiconductor can be reduced to the task of solving the Boltzmann Transport Equation (BTE) [1]. It can be approached in two ways: using Monte Carlo techniques [2][3] or by transforming the BTE into a system of analytically or numerically solvable partial differential equations.

A microscopic, first principle particle Monte Carlo method simulates the motion of a large number of carriers, subject to the action of local electric field and various scattering mechanisms. Each carrier free-flight time and scattering process is selected stochastically according to randomly generated numbers. These random numbers determine the free-flight times, during which the particles accelerate prior to scattering, and the final momentum states after scattering mechanism. The carriers are tracked in real space, momentum space and elapsed flight-time. For each simulation time step Δt, macroscopic observables, such as velocity and energy, can be calculated throughout the device in each space discretisation cell by ensemble averaging [4]. The Monte Carlo approach demands vast computer resources, and is therefore not suitable for a fast, systematic investigation and optimisation of semiconductor devices.

The alternative approach entails the solution of the Boltzmann transport equa-

tion, which can be reduced to a coupled set of macroscopic hydrodynamic equations by taking velocity moments of the BTE [5]. The first, second and third moments correspond, respectively, to statements of conservation of particles (current continuity), conservation of momentum, and conservation of energy. Higher moments of the BTE are decoupled from this set. In the pursuit of semiconductor modeling, these moment equations historically have been formulated using a variety of approximations and assumptions [6][7] but they all go under the name of Drift Diffusion Model (DDM).

With a carefully chosen DDM, it is possible to have a reduction in computer CPU time and cost of several orders of magnitude. Thus, there is great motivation for adopting a simplified DDM approach, instead of a Monte Carlo approach which requires greater complexity and cost even if it may provide, in some instances, a closer approximation to the actual physical phenomena. In any case, the simplified treatments can be used to identify useful device structures before a full computationally intensive simulation is attempted. Furthermore the simplified model utilises several phenomenological parameters derived from Monte Carlo simulations. By fitting the model with these parameters, we build the fundamental aspects of the transport physics into a simpler, tractable transport model.

For the purposes of this thesis we have chosen DDM as our simulation method and this chapter will introduce this numerical method in detail. The drift diffusion equations are outlined in section 2.2. Methods of solving these equations are postponed to section 2.3.

Section 2.4 shows the problem of simulating heterostructures with DDM. There is an intrinsic obstacle due to the continuous nature of the drift-diffusion equations and of the flux of particles in contrast with the presence of hetero-barriers. A different approach based on thermionic theory is described.

Brief details about the implemented numerical solvers are given in section 2.5 together with a description of the boundary conditions we have used. Section 2.6 has been dedicated to the numerical problems that arise in the simulation of wide band gap semiconductor materials.

Quantum wells and tunneling barriers are fundamental parts of semiconductor

devices. Classical approaches, such as the DDM, are no longer valid and alternative methods involving the solution of the Schrödinger equation must be used. Quantum effects can also be observed in simple heterostructures. Heterojunctions introduced by band gap engineering tend to confine electrons in potential wells. These wells, typically on the order of tens of nanometres wide, cause the allowed energy states for carriers to become quantised in the direction perpendicular to the heterojunction. For this reasons in section 2.7 we describe a model for the one-dimensional time-independent Schrödinger equation. Section 2.7.1 outlines the flexibility of the Transfer Matrix Technique (TMT) used to solve the Schrödinger equation. Section 2.7.2 reports how the eigenvalues of Schrödinger equation are found by using the transmission coefficient function. While classical and quantum models are well defined, their interaction is an open problem. Section 2.8 describes our solution of the problem of coupling classical and quantum transport. Finally, section 2.9 introduce few basic concepts about piezoelectric polarization which will be used extensively in chapter 6.

2.2 Drift Diffusion Model

The derivation of the drift diffusion equations can be traced back to Shockley in his 1949 paper. The equations contain a diffusion term which is dependent upon the carrier concentration gradients, and a drift term which is a function of the local field experienced by the carriers. Coupling the Poisson equation with the drift diffusion equations forms 3 partial differential equations that are highly non-linear and this non-linearity makes numerical solutions both complex and unstable. Analytical approaches to solve the drift diffusion equations are centred around simplifying the equations to find a closed form solution [1]. As the speed and memory size of computers have improved, numerical approaches for solving these equations have largely replaced these earlier techniques. The drift diffusion method has been successfully applied to model the d.c. and a.c. characteristics of a whole range of devices from junction diodes through to heterostructure transistors and beyond. This approach can be readily applied in one, two and even three dimensions without prohibitive simulation times or memory usage.

2.2.1 Continuity Equations

The phenomenological hydrodynamic equations governing the physical model for carrier transport in semiconductor materials are derived from the first two moments of the BTE [8], and express the current continuity and the particle current relationships through the two parabolic partial differential equations (PDEs)

$$\frac{\partial n}{\partial t} = \frac{1}{q}\nabla \cdot \mathbf{J}_n + G - R \tag{2.1}$$

$$\frac{\partial p}{\partial t} = -\frac{1}{q}\nabla \cdot \mathbf{J}_p + G - R \tag{2.2}$$

where

$$\mathbf{J}_n = -q\mathrm{v}_n(\mathcal{E})n + qD_n\nabla n \tag{2.3}$$

$$\mathbf{J}_p = q\mathrm{v}_p(\mathcal{E})p - qD_p\nabla p. \tag{2.4}$$

Here n and p are respectively the electron and hole densities, $\mathbf{J}_n$ and $\mathbf{J}_p$ are electron and hole current densities, q is the electron charge, μ_n and μ_p are electron and hole mobility, D_n and D_p are electron and hole diffusion coefficients, $\mathcal{E}$ is the electric field. G and R are respectively generation and recombination processes. A detailed description of these terms will be given in section 2.2.5. The drift velocities $\mathrm{v}_n(\mathcal{E})$ and $\mathrm{v}_p(\mathcal{E})$ are obtained from Monte Carlo simulations.

2.2.2 Poisson's Equation

The electric field, $\mathcal{E}$, is determined from the Poisson equation, where the displacement field $\mathbf{D}$ is given by

$$\nabla \cdot \mathbf{D} = \rho \tag{2.5}$$

and

$$\mathbf{D} = \epsilon\mathcal{E} + \mathbf{P} \tag{2.6}$$

where $\epsilon = \epsilon_0\epsilon_r$ is the dielectric permittivity depending on particular material,

expressed as the product of the dielectric constant in vacuum ϵ_0 and the relative dielectric coefficient ϵ_r for the specific material. $\mathbf{P}$ is the polarization of the material. The charge density ρ is calculated from the free electron n and hole p densities, the density of ionised dopant atoms N_D^+ (donors) and N_A^- (acceptors), and the local distributions of the bound carries of the quantum well n_{QW} and p_{QW}:

$$\rho = q\left(p - n + N_D^+ - N_A^- + p_{QW} - n_{QW}\right). \tag{2.7}$$

The scalar potential, ψ, is linked to the electric field by

$$\mathcal{E} = -\nabla\psi. \tag{2.8}$$

The general form of equations 2.5, 2.6 and 2.7 is suitable for modeling heterostructure devices and quantum wells. The inclusion of the polarization $\mathbf{P}$ in equation 2.6 is necessary for the simulation of piezoelectric semiconductors as reported in chapter 5.

The ionised donor and acceptor impurity concentrations N_D^+ and N_A^- are given by:

$$N_D^+ = \frac{N_D}{1 + \frac{n}{N_C} g_D \exp\left(\frac{\Delta E_D}{kT}\right)} \tag{2.9}$$

$$N_A^- = \frac{N_A}{1 + \frac{p}{N_V} g_A \exp\left(\frac{\Delta E_A}{k_B T}\right)} \tag{2.10}$$

where N_D, N_A are the donor and acceptor doping concentrations, g_D, g_A are the ground state degeneracy factors of the donor and acceptor level and ΔE_D, ΔE_A are the donor and acceptor activation energies. k_B is the Boltzmann constant and T is the temperature in Kelvin. N_C and N_V are the effective densities of states in the conduction and valence bands respectively. In the effective mass approximation the densities of states are given by

$$N_C = 2\left(\frac{2\pi m_e k_B T}{h^2}\right)^{\frac{3}{2}} \tag{2.11}$$

$$N_V = 2\left(\frac{2\pi m_h k_B T}{h^2}\right)^{\frac{3}{2}}. \tag{2.12}$$

Here m_e and m_h are respectively the effective masses for electrons and holes and h is the Planck constant.

2.2.3 Carrier Density

In equilibrium, the density of electrons in a semiconductor is given by

$$n = \int_{E_C}^{\infty} g(E)f(E)\mathrm{d}E. \tag{2.13}$$

where E_C is the energy at the bottom of the conduction band and $f(E)$ is the Fermi-Dirac distribution function. The density of states $g(E)$ has been approximated by the density near the bottom of the conduction band (where the band is parabolic) for low-enough carrier densities and temperatures:

$$g(E) = M_c \frac{\sqrt{2}}{\pi^2} \frac{(E - E_C)^{1/2}}{\hbar^3} (m_e)^{3/2} \tag{2.14}$$

where M_c is the number of equivalent minima in the conduction band. Substituting equation 2.14 and $f(E) = 1/[1 + \exp((E - E_n)/k_B T)]$ in equation 2.13, we obtain

$$n = N_C F(E_n - E_C) \tag{2.15}$$

where $F = 2f_{1/2}/\sqrt{(\pi)}$, $f_{1/2}$ is the Fermi integral defined by

$$f_j(x) = \frac{1}{j!}\int_0^\infty \frac{t^j}{\exp(t-x)+1}\mathrm{d}t \tag{2.16}$$

and E_n is the quasi-Fermi level for electrons. Note that equation 2.15 can be used to define E_n. A similar expression can be obtained for holes in the valence band:

$$p = N_V F(E_V - E_p) \tag{2.17}$$

where now E_V is the top of the valence band and E_p is the quasi-Fermi level for holes.

In calculations, we find it convenient to use the approximate analytical formulae for F proposed in [9]

$$F(x) \approx (e^{-x} + \xi(x))^{-1} \tag{2.18}$$

$$\xi(x) = \frac{3}{4}\sqrt{\pi}(\mu(x))^{3/8} \tag{2.19}$$

$$\mu(x) = x^4 + 50 + 33.6x\left[1 - 0.68\exp[-0.17(x+1)^2]\right]. \tag{2.20}$$

This analytical expression not only approximates the Fermi integral to within 0.4% of error in all ranges, but it also allows an analytical expression for the derivative of the semiconductor equations when Newton method is used for linearisation (see section 2.3).

Equations 2.15 and 2.17 constitute the Fermi-Dirac statistics description of the density of carriers in a semiconductor. When the Fermi levels lie several k_BT below the conduction band for electrons and above the valence band for holes, the Fermi integral F reduces to a simple exponential function and equations 2.15 and 2.17 can be rewritten as

$$n = N_C \exp\left(-\frac{E_C - E_F}{k_B T}\right) \tag{2.21}$$

$$p = N_V \exp\left(-\frac{E_F - E_V}{k_B T}\right) \tag{2.22}$$

where N_C and N_V are the effective densities of states in the conduction and valence bands. Equations 2.21 and 2.22 are termed Boltzmann statistics.

Using Boltzmann statistics it is possible to directly relate the carrier concentrations n and p to the electrostatic potential ψ by using

$$n = uN_C \exp\left(\frac{q\psi + \chi_c}{k_B T}\right) \tag{2.23}$$

$$p = vN_V \exp\left(-\frac{q\psi + \chi_c + E_g}{k_B T}\right) \tag{2.24}$$

where

$$E_C = -q(\psi + \chi) \tag{2.25}$$

$$E_V = -q(\psi + \chi + E_g) \tag{2.26}$$

$$E_n = -q\phi_n \tag{2.27}$$

$$E_p = -q\phi_p \tag{2.28}$$

$$u = \exp\left(-\frac{q\phi_n}{k_B T}\right) \tag{2.29}$$

$$v = \exp\left(\frac{q\phi_p}{k_B T}\right). \tag{2.30}$$

Here χ_c is the electron affinity and E_g the energy gap.

For heavily doped semiconductor devices or for heterostructures the use of Boltzmann statistics gives rise to significant errors and Fermi-Dirac statistics has to be used. Yoshida [10] has shown that Fermi-Dirac statistics give a reasonable approximation for the electron concentration at an AlGaAs heterointerface compared with an exact quantum mechanical calculation, whereas Boltzmann statistics do not. Li et al. [11] have developed a formalism for the convenient use of Fermi-Dirac statistics in the device modeling. Both types of statistics have been implemented in our simulation, using Fermi-Dirac statistics only when required by the complexity of the problem.

2.2.4 Einstein Equations

For nondegenerate semiconductors, the diffusion coefficients, D_n and D_p, can be calculated from the mobilities with the Einstein equations,

$$D_n = \mu_n \frac{k_b T}{q} \tag{2.31}$$

$$D_p = \mu_p \frac{k_b T}{q} \tag{2.32}$$

When Boltzmann statistics are used, the Einstein relation follows from the Boltzmann transport equation [12]. This result cannot be used for degenerate semiconductors, since equations 2.31 and 2.32 are based on the Boltzmann approximation [13][14]. When Fermi-Dirac statistics are used, however, one usually does not start from the Boltzmann transport equation since it is rather complicated. Alternatively, one can start from the phenomenological expression for the electron current given by equations 2.3 and 2.4. Using Fermi-Dirac statistics for n and requiring $\mathbf{J}_n$ to be zero at equilibrium, one obtains the following modification to the Einstein relation [12]

$$D_n = 2V_T \mu_n f_{1/2}(E_F - E_C)/f_{-1/2}(E_F - E_C) \tag{2.33}$$

where $V_T = k_B T/q$ is the thermal voltage. A similar expression is obtained for the hole diffusion coefficient [12].

However Boltzmann statistics are still widely used in heterostructure device modeling because it is difficult to apply equation 2.33 directly to the semiconductor equations when they are discretised. We will show in section 2.3 that a popular formula for the discretised current continuity equation is in terms of the Bernoulli function ($B(x) = 1/(\exp(x) - 1)$), and the expression for the current is linear in the carrier density. Use of equation 2.33 not only destroys this linear property but also makes the formula for the current extremely complicated. In this work the simpler equations 2.31 and 2.32 have been preferred to 2.33. However, a future development of our model should include the Fermi-Dirac correction of equation 2.33.

2.2.5 Recombination Models

BULK RECOMBINATION

In bulk material we distinguish radiative and non-radiative recombinations. Shockley-Read-Hall (SRH) is the non-radiative recombination mechanisms implemented in our model. As far as radiative recombination is concerned, only spontaneous recombination (direct band gap) has been included. Surface recombination is not considered in the model.

Nonradiative: SRH recombination
Impurity atoms, other than donors and acceptors and some types of crystal defects in a semiconductor, introduce localised energy levels deep in the band gap away from the band edges. These levels act as stepping stones for electrons between the conduction and valence band, making a substantial enhancement in the indirect nonradiative recombination process. Depending on its location in the band gap, a deep level may act as:

- an electron trap,
- a hole trap,
- a recombination centre.

An electron trap has a high probability of capturing a conduction electron and setting it free after some time. Similarly, a hole trap has a high probability of capturing a hole that is subsequently released into the valence band. At a recombination centre the probabilities of electron and hole capture are nearly equal. Thus, electron capture is followed by hole capture, and this results in the elimination of an electron-hole pair. The energy is dissipated in the form of phonons, and the recombination is non-radiative. It can be seen that the most effective recombination centres are those with energy levels near the middle of the energy gap [1]. An energy level closer to the conduction band than to the valence band will preferentially act as an electron trap. Similarly a hole trap will have its energy level closer to the valence band than to the conduction band.

The mechanism of indirect recombination through the deep level centres has been investigated by Hall, Shockley and Read [15]. Considere a semiconductor with a density of recombination centres N_t located at energy E_t. The probability that the centre is occupied by an electron is given by

$$G_n(E_t) = \frac{1}{1 + \exp\left(\frac{E_t - E_n}{k_B T}\right)} \tag{2.34}$$

while the probability that a hole occupies that centre is

$$G_p(E_t) = \frac{1}{1 + \exp\left(\frac{E_p - E_t}{k_B T}\right)} \tag{2.35}$$

and consequently the density of trapped electrons and holes is $n_t = N_t G_n(E_t)$ and $p_t = N_t G_p(E_t)$ respectively, and n_t and p_t have to be included in equation 2.7 as they represent a fixed charge affecting the potential through Poisson equation 2.5. The general form of the SRH recombination rate is given by

$$R = \frac{np - n_i^2}{\tau_p(n + n_1) + \tau_n(p + p_1)} \tag{2.36}$$

Here n_i is the carrier concentration at thermal equilibrium of an intrinsic semiconductor, τ_n and τ_p are the lifetime constants, n_1 and p_1 represent the equilibrium electron and hole concentrations that would result when the Fermi level lies at the trap level E_t. Equations 2.15 and 2.17 can be used to evaluate n_1 and p_1 by substituting E_C and E_V with E_t. Equation 2.36 guarantees that recombination is zero at thermal equilibrium. A problem here is to find a value for the radiative lifetimes τ_n and τ_p. These parameters are material and even sample dependent. In the original formulation [15] the lifetime coefficients were expressed as

$$\tau_n = \frac{1}{\sigma_n v_{th} N_t} \qquad \tau_p = \frac{1}{\sigma_p v_{th} N_t}. \tag{2.37}$$

Several values of the cross-sectional areas for electrons σ_n and holes σ_p are given in the literature [1]. However, in our simulations, the lifetime coefficients have been treated as adjustable parameters.

Radiative: Spontaneous recombination

The spontaneous recombination can be approximated by

$$R_{sp} = B_{eff}(n - n_0)(p - p_0) \tag{2.38}$$

where B_{eff}=1·10^{-10} cm^3/s [14]. An improved model includes the case in which the material is degenerate

$$R_{sp} = (B_{eff} - B_1 n)(n - n_0)(p - p_0) \tag{2.39}$$

where B_1/B_{eff}=2·10^{-19} cm^3 for T=300 K. These models are limited as they cannot predict how the emission is changed if the material or temperature is varied. Furthermore, B_{eff} and B_1 are dependent on the material. The reported value were used for GaAs. Large variations of those parameters are possible. A better approximation calculated from first principles is proposed by Grinberg [16], including temperature dependence and composition dependence.

RECOMBINATION IN A QUANTUM WELL

Nonradiative: Traps and Shockley-Read-Hall recombination

Whilst no examples have been found, there is no reason why Shockley-Read-Hall recombination should not be used for a quantum well. A problem is to find a suitable lifetime that is likely to be sample dependent. One can expect that the trap density should be higher compared to bulk due to interface roughness. In that case the lifetime should be shorter than the bulk lifetime. The Shockley-Read-Hall recombination is implemented in the simulator as a separate recombination time constant for quantum wells.

Radiative: Spontaneous emission

It is important to include the effect of degeneracy for quantum wells as the conduction ground state is easily filled to degeneracy. To be able to compare results from time resolved luminescence the absolute value of the spontaneous recombination is important. A good model is given by the following equations [16]:

$$R_{sp} = \frac{1}{E_g} k_{sp} \frac{2}{3} |M|^2 \rho_{opt} k_\rho (k_B T)^{3/2} \frac{(n - n_0)(p - p_0)}{N_{Cqw} N_{Vqw} \left[1 + \frac{n+p-\sqrt{np}}{N^*}\right]}. \tag{2.40}$$

where n_0 and p_0 are the thermal equilibrium densities and N_{Cqw}, N_{Vqw} are given by

$$N_{Cqw} = \frac{m_e k_B T}{\pi \hbar^2} \tag{2.41}$$

$$N_{Vqw} = \frac{m_h k_B T}{\pi \hbar^2}. \tag{2.42}$$

The other constants involved are

$$N^* = \frac{(m_e + m_h) k_B T}{\pi \hbar^2} \tag{2.43}$$

$$m_r = \frac{m_e m_h}{m_e + m_h} \tag{2.44}$$

$$k_\rho = \frac{m_r}{\pi \hbar^2} \tag{2.45}$$

$$k_{sp} = \frac{\pi q^2 \hbar}{\hat{n} \varepsilon_0 m_0^2} \tag{2.46}$$

where $\hat{n}$ is the refractive index, m_0 is the electron mass and M is the optical matrix element as given in [14][17]. The optical mode density ρ_{opt} can be approximated as

$$\rho_{opt} = \frac{1}{\pi^2} \frac{\hat{n}}{(\hbar c)^3} E_g^2. \tag{2.47}$$

In this model, excitonic recombination is not included. This means that the recombination is underestimated for low carrier densities. Even if excitons are of less importance for higher carrier densities, there is still a Coulomb enhancement of the transition which means that the model will underestimate the spontaneous recombination somewhat.

2.2.6 Temperature-dependent parameters

Inclusion of temperature-dependent parameters is very important for simulation of real devices. For example practical Gunn devices operate at about 3-percent dc-to-microwave conversion efficiency, therefore a large amount of energy in the form of heat has to be dissipated. Even with efficient heat sinking, the operating temperatures of practical devices range from about 50 to 200 $°C$ above room temperature. For a realistic simulation, material properties of the devices at the actual operating temperatures must thus be considered. For Gunn diode simulation, temperature-dependent velocity-field characteristic $v(\mathcal{E})$ of GaAs has been obtained with Monte Carlo simulations. A temperature-dependent direct energy gap has been considered using a simple relation [18]

$$E_g(T) = E_{g0} - C_T \frac{T^2}{T + T_C}. \tag{2.48}$$

The parameters E_{g0}, C_T and T_C are material dependent. In the simulation of GaAs we have used E_{g0}=1.519 eV, C_T=5.408·10^{-4} eV/K and T_C=204 K.

2.2.7 Carrier Momentum effects

The assumptions inherent in this model require that the carrier momentum and energy react instantaneously to the local electric field. For slowly varying spatial and temporal evolution of the electric field, the quasi-static continuity equations and Poisson equation are valid, since the change in field strengths encountered by the moving carriers is small during the time necessary for the momentum and energy distributions to reach the steady-state with the local electric field. For reasonable operating frequencies, device lengths and doping density variations,

the physical model described possesses sufficient dynamic character to simulate time-dependent transient and steady-state problems. Operating frequencies are from d.c. to roughly $1/(6\tau_w)$ (τ_w is the energy relaxation time $\simeq 0.5 \rightarrow 3 \cdot 10^{-12}$ s depending on the material) and sample lengths from one to several hundred microns are considered here.

2.3 Numerical Methods for DDM

The drift and diffusion semiconductor equations 2.1, 2.2, and 2.5 comprise a set of coupled, non-linear partial differential equations.

2.3.1 Poisson Equation

The prototypical partial differential elliptic equation is the Poisson equation 2.5. The solution of the Poisson equation is a typical boundary value problem, where the aim is to find a single time independent function $\psi(x, y, z)$ which satisfies the equation within some (x, y, z) region of interest, and which has some desired behaviour on the boundary of the region of interest. In contrast to the continuity equations (see 2.3.2), stability is relatively easy to achieve for the Poisson equation. Thus, the efficiency of the algorithms, both in computational load and storage requirements, becomes the principal concern. Because all the conditions on a boundary value problem must be satisfied simultaneously, these problems usually boil down, at least conceptually, to the solution of large numbers of simultaneous algebraic equations. One way to obtain that is by using the finite difference scheme.

Finite-difference method

The finite-difference method requires the domain to be replaced by a grid, see figure 2.1 for a two-dimensional simulation.

At each grid point each term in the partial differential is replaced by a difference

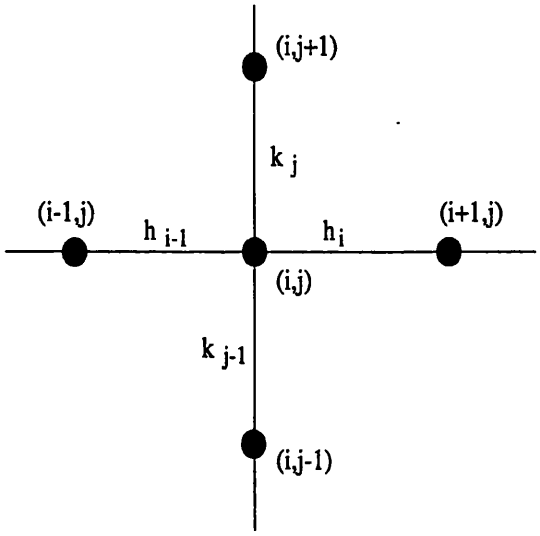

Figure 2.1: Two-dimensional variable mesh.

formula which may include the values of variables at that and neighbouring grid points. This can be accomplished using finite difference approximations to the differential operators:

$$\frac{\partial^2\psi}{\partial x^2} = \frac{2}{h_i + h_{i-1}}\left[\frac{\psi_{i+1}}{h_i} - \left(\frac{1}{h_i} + \frac{1}{h_{i-1}}\right)\psi_i + \frac{\psi_{i-1}}{h_{i-1}}\right] \tag{2.49}$$

$$\frac{\partial\psi}{\partial x} = 2\frac{\psi_{i+1} - \psi_{i-1}}{h_i + h_{i-1}}. \tag{2.50}$$

The continuous derivatives of the semiconductor equations are replaced by the discretised finite-difference approximations. This process represents, together with the choice of variables, a key issue. For a two-dimensional simulation the Poisson equation is usually discretised using a five-point difference approximation, where the potential at a node depends on the potential and charge at the four neighbouring nodes. By substituting the difference formulae 2.49 and 2.50 into equation 2.5, a difference equation is obtained for each point of the mesh:

$$a_{i+1,j}\psi_{i+1,j} + a_{i,j+1}\psi_{i,j+1} + a_{i,j}\psi_{i,j} + a_{i,j-1}\psi_{i,j-1} + a_{i-1,j}\psi_{i-1,j} = b_{i,j} \tag{2.51}$$

$$a_{i,j} = -a_{i+1,j} - a_{i-1,j} - a_{i,j+1} - a_{i,j-1} \tag{2.52}$$

where a and b are known coefficients. The linear system can be solved with many methods as we illustrate in appendix C.

Equations 2.49 and 2.50 are referred to as the central difference approximation but other approximations for the first derivative can be used. The forward-difference formula

$$\frac{\partial \psi}{\partial x} = \frac{\psi_{i+1} - \psi_i}{h_i} \tag{2.53}$$

and the backward-difference formula

$$\frac{\partial \psi}{\partial x} = \frac{\psi_i - \psi_{i-1}}{h_{i-1}} \tag{2.54}$$

have a truncation error higher than the central-difference approximation of equation 2.50. Nevertheless forward and backward-approximations can lead to more stable numerical calculations. This is particularly true for time dependent simulation, where there is a preferential direction for the transmission of information within the equations.

The finite-difference method produces solutions for the physical variable ψ, as a discretised value at specific nodes contained within a mesh superimposed on the solution domain. Non-uniform meshes are considered. A non-uniform mesh allows greater accuracy for the higher valued derivatives and a saving in time and memory in regions of low valued derivatives.

One-dimensional simulation

In a one-dimensional simulation, substituting equations 2.8 and 2.6 in Poisson equation 2.5 the original partial differential equation 2.5 reduces to the an ordinary differential equation (ODE):

$$-\frac{\mathrm{d}}{\mathrm{d}x}\left(\varepsilon \frac{\mathrm{d}\psi}{\mathrm{d}x}\right) + \frac{\mathrm{d}P}{\mathrm{d}x} = \rho(x). \tag{2.55}$$

Equation 2.55 is now required to satisfy boundary conditions at two values of

the independent variable (two point boundary value problem). With a reduction in the size of the linear system involved, it is possible to use easier and faster numerical methods for ODEs. We have implemented two numerical methods for solving equation 2.55. In the *shooting* method we choose values for all of the dependent variables at one boundary, we integrate the ODE by initial values method like Runge-Kutta. We then adjust the free parameter at the starting point to get zero discrepancies at the other boundary point. This approach has shown to be slow for solving Poisson and the continuity equations.

In the *relaxation* methods the differential equations are replaced by finite difference equations on a mesh of points that cover the range of integration. A trial solution consists of values for the dependent variables at each mesh point. The iteration, now called *relaxation*, consists of adjusting all the values on the mesh as to bring them into successively closer agreement with the finite difference equations and, simultaneously, with the boundary conditions. Good initial guesses are the secret of efficient relaxation methods. This works nicely when I/V characteristics of semiconductor devices have to be calculated. Using the previous bias step as initial condition increases speed and efficiency of the solver. This approach has been developed with success by Blades [19] and it is has been used extensively in our solution of the Poisson equation for the one dimensional simulator. For a detailed explanation of the relaxation methods and of SOLVDE (the first order Newton algorithm that implements the relaxation approach) see [20].

2.3.2 Continuity Equations

In contrast to the elliptic Poisson equation, the solution of the parabolic continuity equations 2.1 and 2.2 can present difficulties due to the time-dependence and non-linear velocity-field relationship. Both continuity equations define initial value or Cauchy problems. If the information on n or p is given at some initial time t_0 for all the spatial distribution, then the equations describe how $n(x, y, z, t)$ propagates itself forward in time. The goal of a numerical code should be to track the time evolution with some desired accuracy.

In order to reduce the PDE system to an ODE problem, the differential equations 2.1 and 2.2 are replaced by finite difference equations on a mesh of points that

cover the range of integration as described by the discretisation process in section 2.3.1.

There are two principal approaches for discretising the continuity equations:

- (*i*) utilising the explicit current densities at the nodes (i.e. substituting electron and hole current equations 2.3 and 2.4 in equations 2.1 and 2.2 respectively after being discretised),
- (*ii*) substituting the analytical form of J_n (equation 2.3) and J_p (equation 2.4) in equations 2.1 and 2.2 and subsequently discretising the resulting equations.

Using method (*ii*) a typical expansion of the continuity equation for electrons is

$$\frac{\partial n}{\partial t} = n\frac{\partial(\mu\mathcal{E})}{\partial x} + \mu\mathcal{E}\frac{\partial n}{\partial x} + D_n\frac{\partial^2 n}{\partial x^2} + \frac{\partial D_n}{\partial x}\frac{\partial n}{\partial x} \tag{2.56}$$

The finite-difference equation is then obtained by substituting standard finite difference expansions for the derivatives. Such schemes usually require less computation than method (*i*). However, for devices with high carrier concentrations, or for materials with large diffusion coefficients (eg GaAs or InP), the truncation error can become significant, particularly in the vicinity of depletion regions. This is due to the fact that in method (*ii*) the discretisation is done on a second order partial differential equation while method (*i*) use a two separate steps each requiring the discretisation of a first order partial differential equation. It can be shown [6] that this latter approach gives a lower truncation error.

The expansion and discretisation of current densities can be carried out by assuming that the carrier density varies either in linear or in exponential fashion between adjacent mesh nodes. Scharfetter and Gummel demonstrated that by allowing the carrier density to follow an exponential variation between mesh points, errors due to the discretisation could be minimised [21]. Both linearised and Scharfetter-Gummel methods have been implemented. We have found that for TEDs a linearised scheme is required. Small variations in majority carrier density can not be followed with a Scharfetter-Gummel approach. Method (*ii*)

has been used for the linearised scheme while method (i) has been preferred for the Scharfetter-Gummel scheme.

Time Dependent Simulation

In the discretisation of equations 2.56 for instance, we have several choices for representing the time derivative term. The obvious way is to set

$$\frac{\partial n}{\partial t}|_{i,k} = \frac{n_i^{k+1} - n_i^k}{\Delta t}. \tag{2.57}$$

This is called *forward Euler* differencing. While forward Euler is only accurate to first-order in Δt, it has the advantage that one is able to calculate quantities at time step $k+1$ in terms of only quantities known at time step k. If we use a second-order central expansion for the space derivative evaluated at the time step k the resulting finite-difference approximation is called the Forward Time Centred Space (FTCP) representation or fully explicit. If the central expansion for the space derivative is calculated at at time step $k+1$ the approximation is called Backward Time Centred Space (BTCP) or fully implicit.

A more general scheme is the Crank-Nicholson scheme that can be thought as simply the average of the explicit and implicit approximations. The discretised form of the continuity equation for electrons becomes

$$\frac{\partial n}{\partial t} = \frac{n^{t+1} - n^t}{\Delta t} = \left(\theta \frac{\partial J_n^{t+1}}{\partial x} + (1-\theta)\frac{\partial J_n^t}{\partial x}\right) \tag{2.58}$$

Setting the parameter θ=0 we obtain an explicit scheme, with θ=1 we have a fully implicit scheme. Explicit schemes usually require excessively small time steps to guarantee stability and obtain accurate solutions.

Stability analysis has shown that for FTCP linear explicit solutions of the continuity equation the time step is restricted for stability reasons by

$$\Delta t < min\left(\frac{\Delta x^2}{4D_n}, \frac{2D_n}{v_\infty^2}\right) \tag{2.59}$$

where v_∞ is the high field saturated constant velocity. A fully implicit scheme does not have this limitation even if the small-scale evolution from the initial conditions are inaccurate. The approximations obtained using linear stability analysis theory also provide useful guide lines for non-linear problems. However non-linear instabilities arise when large time steps are used. Therefore it is necessary to constrain the time step to be of the same order of the dielectric relaxation time and the space step should not be significantly larger than a Debye length. A Crank-Nicolson scheme with θ=0.5 is usually adopted, which provides greater accuracy and better convergence properties.

2.4 Heterostructures

Generally, although the drift diffusion approach is very appropriate in the bulk regions of a semiconductor device, it may cause problems at abrupt heterojunctions. Drift and diffusion mechanisms are related to the gradient in the band edges and to the gradient in the carrier densities, respectively. However, at an abrupt heterojunction, these quantities can be discontinuous. The discretised equations, instead, will approximate the infinite gradient with a high value of derivative leading to un-physically high currents. The derivative values will also depend on the mesh used in the discretisation process (see section 2.3), being high for a well refined mesh with grid points spaced closely around the interface [22]. For these reasons an alternative approach to the DDM has to be considered.

Several numerical models for heterostructure devices have been proposed to predict the device characteristics or to optimise the device design [23][24][25]. In most cases, transport equations that have been derived for the continuous band structure are applied to the case with an abrupt heterojunction interface. Most authors adopt a so-called *diffusion model*, where current transport across the heterojunction interface is not considered explicitly and so current continuity equations across the interface are not solved, but quasi-Fermi levels are assumed to be continuous [25]. To illustrate the diffusion model we consider only the

continuity equation for electrons.

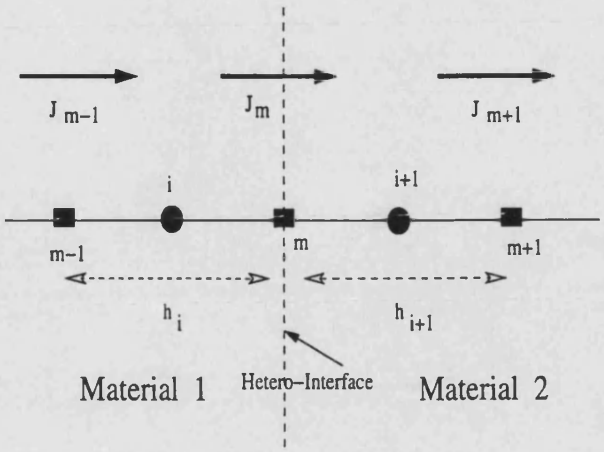

Figure 2.2: Mesh spacing

Using the one-dimensional index notation shown in figure 2.2 and supposing that the interface lies between the mesh point i and $i+1$, where points m, $m-1$ and $m+1$ are fictitious mesh points, discretisation of equation 2.3 around the heterojunction interface can be written as

$$\frac{J_m - J_{m-1}}{h_i} = qR_i \tag{2.60}$$

and

$$\frac{J_{m+1} - J_m}{h_{i+1}} = qR_{i+1} \tag{2.61}$$

for mesh point i and $i+1$, respectively. In these equations, J_{m-1} and J_{m+1} can be formulated using equation 2.3. However, J_m represents the electron current across the heterojunction interface and cannot be expressed in a form similar to equation 2.3. In the diffusion model for heterojunctions, one does not give an expression for J_m explicitly. J_m can be eliminated if we sum equations 2.60 and 2.61, obtaining:

$$J_{m+1} - J_{m-1} = qR_{i+1}h_{i+1} + qR_ih_i \tag{2.62}$$

Since this step has reduced the two-equation system (equations 2.60 and 2.61) to one equation, another independent relation is necessary for a unique solution.

In the diffusion model for heterojunctions, it is assumed that the electron quasi-Fermi level is continuous at the heterojunction interface. This is expressed as

$$E_{Fn,i} = E_{Fn,i+1}, \tag{2.63}$$

As evident from the above treatment, current transport mechanisms across the heterojunction interface are not taken into account explicitly in the diffusion model. Furthermore, equation 2.63 is only an assumption, and it is not a general condition [26].

Thermionic emission theory represents a considerable improvement over the diffusion method [26][27][23][24]. In the present work an expression for J_m has been given considering a transport mechanism across the heterojunction based on thermionic emission. The theory, developed by Wu and Yang [28], takes quantum mechanical effects into account such as different effective masses between the semiconductors, the quantum mechanical transmission into the barrier and the quantum mechanical reflection at the interface. Hence the electron current density at the interface can be expressed as:

$$J_m = q\eta[n_{i+1}v_{R1} - \theta n_i v_{R2}\exp(-\frac{E_B}{kT})] \tag{2.64}$$

with $\theta=\frac{m_2}{m_1}$ where $m_{1(2)}$ is the effective mass in the semiconductor 1(2), and E_B is the barrier in the conduction band. The recombination velocities v_{R1} and v_{R2} in the material 1 and material 2 respectively, are given by

$$v_R = \frac{A^*T^2}{qN_C}. \tag{2.65}$$

A^* is the effective Richardson constant and η is a coefficient that includes quantum mechanical effects [28]. If the effective masses of electrons in both semiconductors are equal (θ=1, η=1, v_{R1}=v_{R2}) and if n_i or n_{i+1} is set equal to the value under equilibrium, equation 2.64 becomes equal to a boundary condition used for metal-semiconductor contacts in device analysis.

Here we apply this model to GaAs/AlGaAs heterojunctions as an example and compare the results with those obtained from the diffusion model. The device structure is a GaAs/$Al_{0.3}Ga_{0.7}As$ diode where each region is 0.5μm in length and n-doped respectively 10^{15} cm^{-3} and 10^{16} cm^{-3}. Figure 2.3 shows current-voltage characteristics calculated using the thermionic model and the diffusion model. Qualitative and quantitative agreement has been achieved with results in the literature [24].

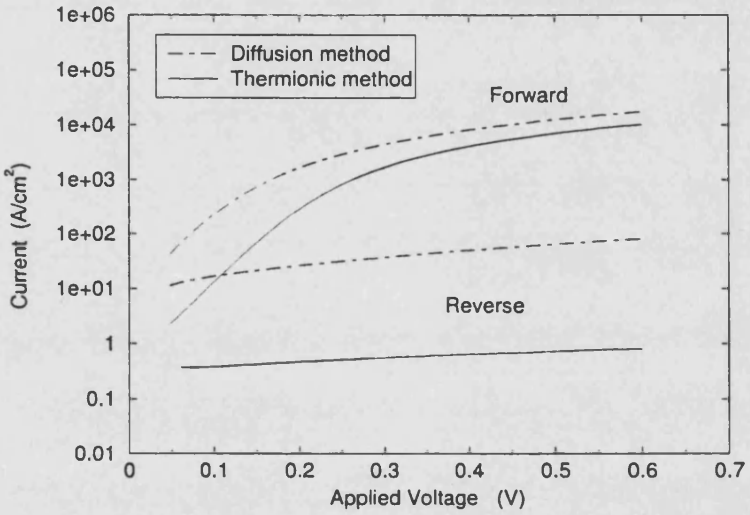

Figure 2.3: Current-voltage characteristics

The results in figure 2.3 indicate that current transport across the heterojunction interface (i.e. the thermionic emission mechanism) limits the overall transport process. Therefore, as was shown before, current-voltage characteristics depend strongly on whether the model includes current transport across the heterojunction interface or not. We conclude that the diffusion model does not lead to a correct description of this problem, most likely because the thermionic emission mechanism becomes important in the case where high barriers for majority carriers are present, as in isotype heterojunctions.

2.5 Solvers and boundary conditions

Simultaneous equations formed by the finite-difference expressions for the Poisson equation and continuity equations can be solved by direct or iterative techniques. We have implemented three different direct methods: tridiagonal direct solver,

LU decomposition solver and Gauss elimination solver. We have also used two iterative techniques: Newton method and successive-over-relaxation (SOR). While direct methods converge quickly for steady-state solutions, iterative methods do not easily converge for time-dependent solutions. For further details see appendix C.

To complete the mathematical description, specifications of initial conditions at the cathode and at the anode are required. Initial conditions are taken from a steady-state solution. Ohmic contacts are usually described by Dirichlet boundary conditions where the potential and carrier concentration are pre-defined at the contacts. In our model the Dirichlet condition is applied to each unknown at the boundaries. A different type of boundary condition (the Neumann condition) can be set to investigate different types of contacts (ohmic/Schottky). The Neumann boundary condition imposes a fixed value for the derivative of the variable. It can be used when a variable has a unknown value at the boundary and can float. Fixing its derivative gives a constant value of the flux of this variable.

2.6 Wide band gap materials

Numerical problems arise when trying to solve the equation 2.1 and equation 2.2 for wide bandgap materials. For example, discretisation of the steady-state one-dimensional electron continuity equation $n\mu_n + D_n\frac{\mathrm{dn}}{\mathrm{dx}} = R - G$ leads to the following

$$a_i u_{i-1} + b_i u_i + c_i u_{i+1} = d_i \tag{2.66}$$

where $u = \exp(-\phi_n/V_T)$ is the Slotboom variable for electrons evaluated in three mesh points, respectively $i-1$, i and $i+1$ as described in section 2.3. The coefficients a_i, b_i, c_i contain a term such as $\exp\{(-\psi - \chi)/V_T\}$ where ψ is the electrostatic potential and χ is the electron affinity. Since the variation of the potential can be of the order of the energy gap, a_i, b_i and c_i present huge exponential variations that cannot be rescaled. Moreover, the recombination coefficient given by d_i contains terms of the order of the intrinsic concentration of carriers which is much smaller than most other semiconductors [1]. Therefore, for the

presence of highly varing matrix coefficients and very tiny ones, direct method solvers are numerically unstable. In this work we have used an iterative indirect method that has a stable though slowly converging solution.

2.7 The Schrödinger Equation

Fabrication of quantum well heterostructures and superlattices has allowed tailoring of optical and electronic responses of the semiconductor materials, resulting in a considerable increase in research activity directed towards the development of novel electronic and optoelectronic devices. It is well known that exact analytical solutions to such problems are available only for for simple potential structures such as the square well, parabolic well, etc., and even in theses structures, in the presence of perturbations such as electric fields, the problem in general cannot be solved exactly.

The bound-state energies of a one dimensional potential well are given by the solutions of the time-independent Schrödinger equation,

$$\frac{\mathrm{d}^2\psi}{\mathrm{d}x^2} + \frac{2m^*}{\hbar^2}[E - V(x)]\psi = 0 \tag{2.67}$$

where m^* is the effective mass and E and ψ respectively are the energy eigenvalue and eigenfunction of the bound state. The potential profile $V(x)$ is obtained from the Poisson equation 2.5 and is an input for the Schrödinger equation solver 2.67. The N energy eigenvalues and N eigenfunctions found by solving 2.67 are then used to calculate the effective density of electrons n_{QW} and holes p_{QW} in the quantum wells with the following relations:

$$n_{QW} = \frac{m_e k_B T}{\pi\hbar^2} \sum_{i=1}^{N} \ln[1 + e^{(E_n - E_i)/k_B T}]|\psi_{n,i}|^2, \tag{2.68}$$

$$p_{QW} = \frac{m_h k_B T}{\pi\hbar^2} \sum_{i=1}^{N} \ln[1 + e^{(E_i - E_p)/k_B T}]|\psi_{p,i}|^2. \tag{2.69}$$

The sum is done over all the energy eigenvalues E_i with $i = 1, 2, \ldots, N$. Each element of the sum is weighted by the density of probability given by the square modulus of the i-th wavefunction ψ_i. Equations 2.68 and 2.69 are then used in the Poisson equation 2.5 and 2.7 in a self-consistent calculation.

Numerical integration of the wave functions in the wells and the surrounding barriers can then be used to calculate the electron-hole overlap integral [38]

$$M_{cv} = \left| \int_{-\infty}^{+\infty} \psi_c(z) \cdot \psi_v^*(z) \mathrm{d}z \right|. \tag{2.70}$$

The oscillator strength is proportional to the square of the overlap integral [29]. In this way one obtains useful information about the optical transition.

2.7.1 Transfer Matrix Technique

Several techniques have been used to solve the Schrödinger equation in single quantum well (SQW) structures in applied electric fields [30][31][32].

1. The variational method has been widely used for studying quantum well structures in the presence of perturbations like electric field [33]. Variational methods are capable of handling nonlinear potentials. However, the method does not give the correct value of the wave function outside the well, and does not predict the lifetime of quasi-bound states. Moreover, the energy level and wave function depend upon the choice of the trial function, which becomes very difficult to guess for structures involving more than one quantum well, and also when the potential energy variation is graded. And, as is well known, the variational method is not practical for states other than the ground state.

2. The Airy function approach [34] is sometimes limited by numerical instabilities since it must be approximated by power series expansions. It also applies only to cases of linear potentials and is not easily adapted to multiple-well structures.

3. Numerical solutions of the time-independent Schrödinger equation using the

Numerov technique [35] are capable of handling a wide variety of potentials. However, the use of numerical forms of the wave functions dictates the subsequent use of numerical methods for calculating matrix elements. The numerical inaccuracy is therefore amplified in any further calculation.

4. The transfer-matrix technique (TMT) [36][37][38] has the potential to overcome all of the previously mentioned difficulties and also to provide a simplified analytical form of the wave function for calculation of matrix elements and other physical quantities of interest. The TMT approximates the potential in equation 2.67 with a series of constant steps [39][40]. The step size required to accurately model the hamiltonian of Schroödinger equation is reported to be less than 10 meV [36]. The memory required to store the wave functions is determined by the number of potential steps necessary to accurately represent the original potential. For slowly varying potentials a huge saving in memory can be obtained. The TMT is therefore a potentially versatile and accurate method for modeling SQW and MQW structures under a variety of external perturbations. A good description of the TMT technique is given by Davies [39] and is summarised in appendix D.

2.7.2 Determination of eigenvalue energies: The Transmission Coefficient Technique

The TMT is not posed as an eigenvalue problem because no information about the eigenvalues of the system can be directly obtained. In order to determine the eigenvalue energies an energy range is scanned and a solution for the wave function at each energy is then used to determine which, if any, of the energies tested correspond to an eigenvalue. There are three characteristic functions involving the wave function $\psi(E)$ which are used to evaluate whether an energy is an eigenvalue; these are the modulus squared of the transmission coefficient $T_C(E)$, the active-region occupation probability $O_P(E)$ and a ratio of the probability between two or more points in the structure $P_R(E)$. For a description of the last two techniques see [40]. We will concentrate our attention on the transmission coefficient technique.

The Transmission coefficient technique locates the eigen-energies by maximising

the $T_C(E)$ function. The full-width half-maximum of a resonance in the modulus squared of the complex transmission coefficient can be fitted to a Lorentzian for well-bound states to directly obtain the lifetime of the state. For realistic structures the transmission coefficient can be calculated only for energies greater than the maximum potential of the boundary regions, i.e. for energies where wave transmission takes place. A technique which enables the transmission coefficient to be calculated at all energies is reported by [37][38] and illustrated in this work.

We consider a potential energy variation $V(x)$ as shown in figure 2.4(a). Such a structure would have a finite number of bound states corresponding to $V_{min} < E < V'$. Although all values of $E > V'$ are allowed, we will have a finite number of quasi-bound states corresponding to $V' < E < V''$. In order to obtain the eigenvalues corresponding to the bound and quasi-bound states, and the lifetimes of the latter, we introduce a region of potential energy which is less than or equal to V_{min} as shown in figure 2.4(b); we also choose the origin at the point where we introduce the potential.

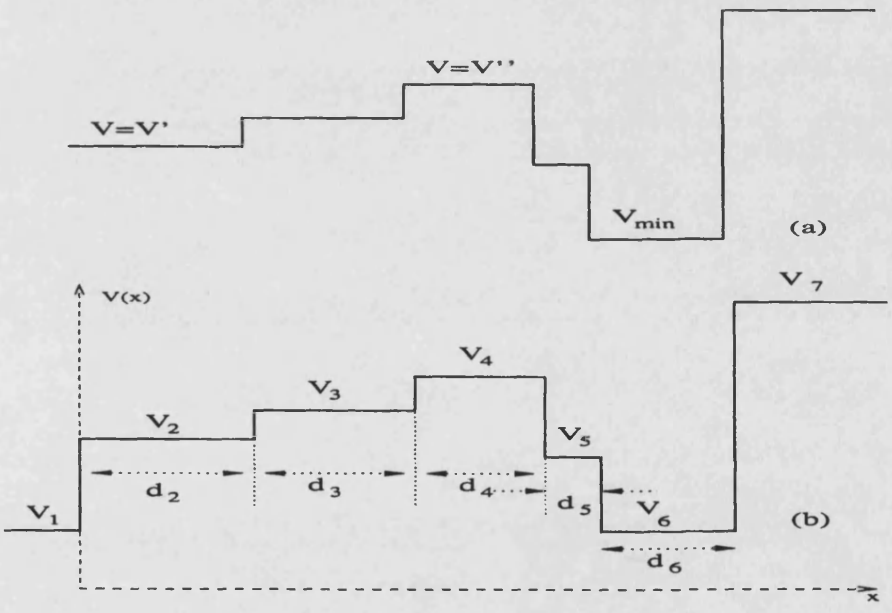

Figure 2.4: Potential energy step variation

Referring to figure 2.4(b), the solution of the Schrödinger equation in each region may be written in the form

$$u_i = u_i^+ e^{-i\Delta_i} e^{ik_i x} + u_i^- e^{i\Delta_i} e^{-ik_i x} \tag{2.71}$$

where Δ_1=Δ_2=0 and

$$\Delta_i = k_i(d_2 + d_3 + \cdots + d_{i-1}) \qquad i = 3, 4, \cdots \tag{2.72}$$

$$k_i = \left[\frac{2m}{\hbar^2}(E - V_i)\right]^{1/2}. \tag{2.73}$$

u_i^+ and u_i^- represent the amplitudes of the waves propagating along the $+x$ and $-x$ directions, respectively. Using the T-matrix technique (see appendix D), continuity of u and $\mathrm{d}u/\mathrm{d}x$ at each interface gives us

$$\begin{pmatrix} u_1^+ \\ u_1^- \end{pmatrix} = T_1 \begin{pmatrix} u_2^+ \\ u_2^- \end{pmatrix} = T_1 T_2 \begin{pmatrix} u_3^+ \\ u_3^- \end{pmatrix} = \cdots = T_1 T_2 \cdots T_{N-1} \begin{pmatrix} u_N^+ \\ u_N^- \end{pmatrix} \tag{2.74}$$

where the T-matrix T_i is given by

$$T_i = \frac{1}{t_i} \begin{pmatrix} e^{-i\delta_i} & r_i e^{-i\delta_i} \\ r_i e^{i\delta_i} & e^{i\delta_i} \end{pmatrix} \quad \delta_i = k_i d_i \quad r_i = \frac{k_i - k_{i+1}}{k_i + k_{i+1}} \quad t_i = \frac{2k_i}{k_i + k_{i+1}} \tag{2.75}$$

with $i = 1, 2, \cdots, N-1$, N being the total number of regions. In the last region u_N^- would represent an exponentially amplifying solution and should vanish. Setting $u_N^- = 0$, we can calculate the amplitude of the wave function in any region in terms of u_1^+. For the potential energy variation shown in figure 2.4(b), if we calculate the transmission coefficient $|u_6^+/u_1^+|^2$ as a function of E, $T_C(E)$, we would obtain Lorentzian peaks corresponding to each bound state and quasi-bound state of the form [37][41]

$$T_C(E) = |\frac{u_6^+}{u_1^+}|^2 = \frac{\sigma}{(E - E_b')^2 + \Gamma^2} \tag{2.76}$$

where $E = E_b'$ represents the position of the peak and 2Γ the full width at half maximum (FWHM). In order to calculate the transmission coefficient a region of potential V_1 was fictitiously introduced as shown in figure 2.4(b). We now

remove this approximation by just moving that region to infinity, obtaining back the original band profile of figure 2.4(a). As the distance d_2 is increased, E_b' would tend to the value E_b which represents the energy of the bound (or quasi-bound) state; the quantity Γ would tend to zero for bound states and to a nonzero value for quasi-bound states. The limiting values of $\hbar/2\Gamma$ is the lifetime of the quasi-bound state. Thus in order to obtain the energies and lifetimes of the various states of an arbitrarily graded potential well, we adopt the following procedure.

1. We first consider a range of values of potential energy that includes the minimum value inside the potential well, beyond the well.
2. We then calculate the ratio $|u_j^+/u_1^+|^2$ as a function of E, where u_j^+ corresponds to the j-th region.
3. We fit a Lorentzian of the form given by equation 2.76 and obtain E_b' and Γ.
4. We let the region move to ∞ and obtain the limiting values of E_b' and Γ, which give the eigenvalue and lifetime of the bound/quasi-bound state. In practice, the range is moved sufficiently far for E_b' and Γ to converge to the required accuracy.

An arbitrary potential energy variation can be represented by a large number of steps and solved in a similar manner. However, if there is more than one well in the structure, we have to calculate the ratio $|u_j^+/u_1^+|^2$ inside each of the wells so that all energy states are obtained.

Application to Quantum Well Structures

In order to show the applicability of the proposed technique to quantum well structures we consider a conduction electron in a finite GaN-$In_{0.467}Ga_{0.533}N$ single quantum well of depth V_0=0.392 eV (which corresponds to a 56-44% splitting between the conduction and valence band) and of width 5 nm with an effective mass m=0.2m_0. The dispersion equation for a general square quantum well V_0 deep and $2a$ wide is

$$\tan(\alpha a) = \frac{\gamma}{\alpha}, \tag{2.77}$$

for even eigenvalues and

$$-\cot(\alpha a) = \frac{\gamma}{\alpha}, \tag{2.78}$$

for odd eigenvalues, where

$$\alpha^2 = \frac{2mE}{\hbar^2}, \quad \gamma^2 = \frac{2m(V_0 - E)}{\hbar^2}. \tag{2.79}$$

If we solve equations 2.77 and 2.78 for the system previously described and if we consider the bottom of the well as reference energy level (E=0), we obtain three eigenvalues E_1=0.0457 eV, E_2=0.177 eV and E_3=0.359 eV. In figure 2.5 we present the transmission energy versus the energy for the same quantum well system as described above. The three peaks correspond roughly to the three eigenvalues. We note that while in the simulation of figure 2.5 the heterostructure model has been fully included, an homostructure is considered in the analytical solution of equations 2.77 and 2.78.

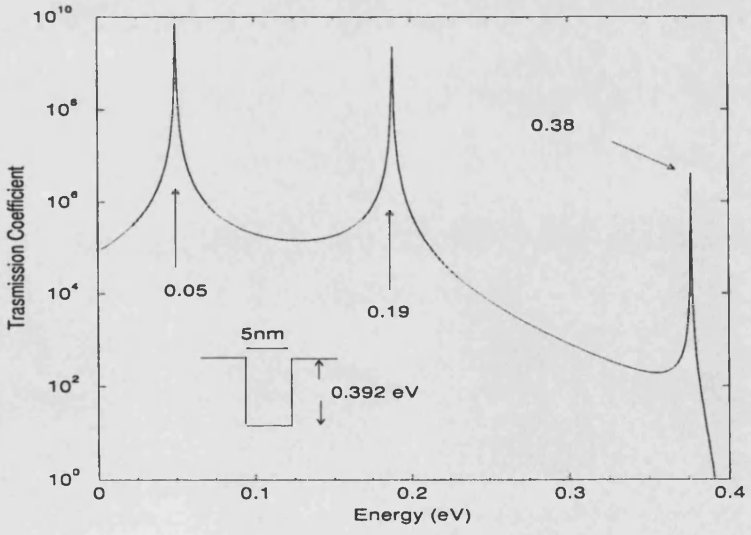

Figure 2.5: Transmission coefficient function for a GaN-$In_{0.467}Ga_{0.533}N$ single quantum well. Three lorentzian peaks are visible indicating the position of the eigenvalues.

2.8 Classical and quantum models

The characteristics of a single or multiple quantum well device depend on carrier transport and the resulting local carrier concentrations throughout the entire device structure. The rate at which carriers can reach the active region largely depends on the drift of carriers through the optical confinement region. Inevitably, some carriers that are incident on the active region will be either be involved in non-radiative processes or will not be captured in the quantum well and recombine at the contacts (dark current). The dynamics of carriers that supply optical emission and those that are wasted as dark current poses a complicated problem involving carrier capture in the quantum well and transport in the surrounding classical regions. Therefore, the difficult question is how to couple the classical carriers in the bulk region to the quantised carriers of the quantum wells [42]. There have been many attempts to solve this problem. A complex full quantum mechanical approach [44] has been substituted by a simpler Wigner transform method as proposed in [45] or by a Wigner-function formalism in [46]. This approach however is very complicated and computationally intensive.

We have used a simpler picture, where the carriers drift and diffuse to a region close to the quantum wells. Over the quantum well, the carriers continue to move with drift/diffusion, but some of the carriers can be captured into the quantum wells [47]. The exact choice of region where carriers can be captured is not critical. The simplest approach, and the one we have used in our simulations, is just to use the quantum well width. Kolbas et al. [48] experimentally indicate that one instead should use the wavefunction extension or more precise the full-width at half maximum of $|\psi|^2$. However for normal sized quantum wells (under 10 nm), the difference is small.

To calculate the capture current, a thermionic model is used in our simulation [49] as it can handle capture in a physically consistent way for situations far from thermal equilibrium. Simpler models using rate equations for capture give problems as they do not guarantee zero current at thermal equilibrium.

Thermionic model for capture in quantum well

In the following discussion, only the conduction band is considered. The current is the electron current and the charges are due to electrons in conduction band. The valence band is treated in an analogous way. The captured current can be

written as

$$J_{QW} = -C\left[e^{\left(\frac{E_{F,B}-V_b+\psi}{k_BT}\right)} - e^{\left(\frac{E_{F,QW}-V_b+\psi}{k_BT}\right)}\right]. \tag{2.80}$$

Here $E_{F,B}$ and $E_{F,QW}$ are the quasi Fermi levels of the barrier and the quantum well, respectively, V_b is the barrier potential and ψ is the electrostatic potential. This expression guarantees zero capture current for thermal equilibrium ($E_{F,B} = E_{F,QW}$). The problem is to find the value of the coupling constant C. One cannot expect the coupling constant to be independent of the carrier densities, since with increasing carrier densities, carrier-carrier scattering is increased but scattering from defects is reduced due to screening of the Coulomb field. Lacking better models we will use a carrier independent value for the coupling constant. An expression for C has been found in the literature [14]

$$C = 2\frac{q(k_BT/\pi)^{3/2}\sqrt{m_e/2}}{L\hbar^2} \tag{2.81}$$

where L is the well width. Equation 2.81 has been used as starting value of our simulations. Investigations about the dependence of our results on C need to be carried out more systematically. When results differed significantly from experimental results the constant C was chosen as fitting parameter.

Tunneling

When modelling multi-quantum well structures, electron and hole tunneling between adjacent wells has to be considered. The tunneling current is given by [39]

$$I = \frac{2q}{h}\sum_{E_C}^{\infty}[f(E, E_{F,qw1}) - f(E, E_{F,qw2})]\,T_C(E)\mathrm{d}E \tag{2.82}$$

where T_C is the transmission coefficient of the barrier between the two quantum wells ($qw1$,$qw2$).

2.9 Piezoelectric field

In the absence of external electric fields, the total macroscopic polarization P of a solid, included in equation (2.5), is the sum of the spontaneous polarization P_{sp} in the equilibrium structure, and of the strain-induced or piezoelectric polarization P_{pz}. In the linear regime assumed here, P_{pz} is related to the strain ϵ by (Voigt notation)

$$P_{pz,i} = \sum_j e_{ij}\epsilon_j \tag{2.83}$$

which defines the components of the piezoelectric tensor e_{ij}. III-V nitrides have a wurtzite structure. This structure has the highest symmetry compatible with the existence of a spontaneous polarization, and its piezoelectric tensor has three non-vanishing independent components. We restrict our model to polarizations along the (0001) axis as this is the direction along which bulk material and superlattices are grown. Hence $P_{pz,3}$ is expressed in terms of the piezoelectric coefficients e_{33} and e_{31} as [51]

$$P_{pz,3} = e_{33}\epsilon_3 + e_{31}(\epsilon_1 + \epsilon_2) \tag{2.84}$$

where a_0 and c_0 are the equilibrium values of the lattice constants of a relaxed InGaN alloy with an In content of x; $\epsilon_3=(c-c_0)/c_0$ is the strain along the c axis, and the in-plane strain $\epsilon_1=\epsilon_2=(a-a_0)/a_0$ is assumed to be isotropic. In the case of wurtzite lattice grown along the (0001) axis with an isotropic in-plane strain and no applied stress on the c-direction, we define the Poisson's ratio ν using the following equation

$$\frac{\nu}{1-\nu} = -\frac{\epsilon_3}{\epsilon_{1(2)}} = -\frac{C_{13}}{C_{33}}, \tag{2.85}$$

where C_{13} and C_{33} are components of the elastic tensor of the lattice.

Experimentally x is obtained from measurements of the lattice constants. Here

we assume the validity of Vegard's Law. Then, in the case of fully relaxed layers, it is sufficient to measure one of the lattice constants, usually c, to calculate x. For strained layers, this is no longer true as the change of the lattice constants due to biaxial stress must be considered. To divide the influence of strain and composition on the lattice constants, both lattice parameters, c and a, have to be determined. In strained films c and a are connected via the Poisson's ratio,

$$[c - c_0(x)] + \frac{\nu(x)}{(1-\nu(x))}\frac{c_0(x)}{a_0(x)}[a - a_0(x)] = 0 \tag{2.86}$$

A complete solution of equation(2.86) can be found in [52]. For an estimate of x in a strained or partially relaxed InGaN layer, a fixed Poisson's ratio can be assumed, so equation(2.86) can be simplified as following

$$x = \left(\frac{1-\nu}{1+\nu}\right)\left(\frac{c - c_{GaN}}{c_{InN} - c_{GaN}}\right) + \left(\frac{2\nu}{1+\nu}\right)\left(\frac{a - a_{GaN}}{a_{InN} - a_{GaN}}\right). \tag{2.87}$$

Therefore, predicting the effect of strain on electronic properties requires the elastic constants [53]. Comparisons of the available measured values with results obtained in previous theoretical studies reveal clear discrepancies between the different sets of elastic constants (C_{13},C_{33}) for GaN and InN [54] as summarised in table 5.2.3 of chapter 5. Furthermore, piezoelectric coefficients e_{33} and e_{31} present the same difficulty [51][55][56][57][60].

Recent experiments [61] have obtained the bowing parameter for InGaN as 2.5 eV. They find that the main reason of scatter in published values of the InGaN bowing parameter is the uncertainty in ν and show that after recalculation to the same value of ν=0.17, most published data yield the same results within experimental uncertainty. This value has been used in our simulations.

Substituting equations 2.86 and 2.85 in equation 2.84 we obtain

$$P_{pz} = \left(-e_{33}\frac{C_{13}}{C_{33}} + 2e_{31}\right)\epsilon_1. \tag{2.88}$$

In order to compare wurtzite nitrides with other semiconductors, the quasi-cubic

approximation (see section 6.2.7 in chapter 6) will be assumed in some cases. According to this approximation we can rewrite the expression of the piezoelectric polarization using the typical cubic lattice piezoelectric coefficient e_{14}

$$P_{pz} = \frac{2}{\sqrt{3}} \left(-e_{14} \frac{C_{13}}{C_{33}} - e_{31} \right) \epsilon_1. \tag{2.89}$$

2.10 Conclusion

We have presented a model based on a simple hydrodynamic transport equation coupled with self-consistent solution of Poisson and Schrödinger equations. We have outlined numerical methods to solve these equations in the case of two- or one-dimensional simulation or even time dependent. This model can be based on several phenomenological parameters derived from Monte Carlo particle simulations, and therefore the fundamental aspects of the semiconductor transport physics are built into simpler, tractable questions. Since the computer program is quite general, it should be noted that transport parameters from semiconductor materials other than GaAs, such as Si and InP, can be used as well. We have added to the model the possibility of simulation of wide band gap materials. Characteristics typical of III-nitrides such as strain and piezoelectric polarization are built in the model. Several orders of magnitude reduction in computer costs and resources over the Monte Carlo method can be realised permitting efficient and systematic study of device characteristics. Furthermore qualitative insights into important internal physical processes of carrier transport phenomena can be gained.

References

[1] S. Sze, *Physics of Semiconductors*, John Wiley and Sons, New York (1981).

[2] R. J. Mills, PhD thesis, University of East Anglia (UK) 1994.

[3] P. J. Briggs, PhD thesis, University of East Anglia (UK) 1997.

[4] C. Jacoboni and P. Lugli, *The Monte Carlo Method for Semiconductor Device simulation*, Springer-Verlag Wien, New York, 1989.

[5] K. Hess, *Advanced theory of semiconductor devices*, Prentice-Hall (1988).

[6] C.M. Snowden, *Introduction to Semiconductor Device Modelling*, World Scientific, ISBN 9971-50-142-2, 1986.

[7] M. Reiser, *Large-Scale Numerical Simulation in Semiconductor Device Modelling*, Computer Methods in Applied Mechanics and Engineering, **1**, 17 (1972).

[8] K. F. Brennan, *The Physics of Semiconductors with Application to Optoelectronic Devices*, Cambridge University Press, Cambridge (1999).

[9] D. Bednarczyk and J. Bednarczyk, *Phys. Lett.*, **64A**, 409 (1978).

[10] J Yoshida, *IEEE Trans. Electron Devices*, **ED-33**, 154 (1994).

[11] Z-M. Li, S. P. McAlister and C. M. Hurd, *Semicond. Sci. Technol.*, **5**, 408 (1990).

[12] R. A. Smith, *Semiconductors*, Wiley 2nd ed., New York, 1978.

[13] A.H. Marshak and C.M. van Vliet, *Proceeding of the IEEE*, **72**, 148 (1984).

[14] K. Fröjdh, *Carrier Transport Effects in Semiconductor Heterostructures for Optical Applications*, PhD Thesis, Royal Institute of Technology, Sweden 1996.

[15] W. Shockley and W. T. Read, *Phys. Rev. B*, **87**, 835 (1952).

[16] A.A. Grinberg, *IEEE J. Quantum Electron.*, **30**, 1151 (1994).

[17] A. Niwa, T. Ohtoshi and T. Kuroda, *Japanese Journal of Applied Physics*, **35**, L599 (1996).

[18] J. S. Blakemore, *J. Appl. Phys.*, **53**, R123 (1982).

[19] C. D. J. Blades, PhD Thesis, University of Bath 2000.

[20] W. H. Press et al, *Numerical Recipes*, Cambridge University Press.

[21] H. K. Gummel, *IEEE Trans. Elec. Dev.*, **11**, 455 (1964).

[22] M. Grupen, U. Ravaioli, A. Galick, K. Hess and T. Kerkhoven, *SPIE*, **2146**, 133 (1994).

[23] G.B. Tait and C.R. Westgate, *IEEE Transactions on Electron Devices*, **38**, 1262 (1991).

[24] K. Horio and H. Yanai, *IEEE Transactions on Electron Devices* , **37**, 1093 (1990).

[25] M. S. Lundstrom and R. J. Schuelke, *IEEE Transactions on Electron Devices*, **30**, 1151 (1983).

[26] M. J. Adams and A. Nussbaum, *Solid-State Electron.*, **22**, 783 (1979).

[27] K. Yang, J. R. East and G. I. Haddad *Solid-State Electron.*, **36**, 321 (1993).

[28] C. M. Wu and E. S. Yang, *Solid-State Electron.*, **22**, 241 (1979).

[29] D. A. B. Miller, J. S. Wiener and D. S. Chemla, *IEEE J Quantum Electron*, **22**, 1816 (1986).

[30] D. A. B. Miller, D. S. Chemla, T. C. Damen, A. C. Gossard, W. Wiegmann, T. H. Wood and C. A. Burrus, *Phys. Rev. B.*, **32**, 1043 (1985).

[31] P. W. D. Mcllroy, *J. Appl. Phys.*, **59**, 3532 (1986).

[32] J. Singh, *Appl. Phys. Lett.*, **48**, 434 (1986).

[33] G. Bastard, E. E. Mendez, L. L. Chang and L. Esaki, *Phys. Rev. B.*, **28**, 3241 (1983).

[34] T. Hiroshima and R. Lang, *Appl. Phys. Lett.*, **49**, 639 (1986).

[35] P. C. Chow, *Amer. J. Phys.*, 730 (1972).

[36] A. Harwit and J. S. Harris, *J. Appl. Phys.*, **60**, 3211 (1986).

[37] A. K. Ghatak, K. Thyagarajan and M. R. Shenoy, *IEEE J. of Quantum Electronics*, **24**, 1524 (1988).

[38] P. J. Stevens, M. Whitehead, G. Parry and K. Woodbridge, *IEEE J. Quantum Electron.*, **24**, 2007 (1988).

[39] J. H. Davies, *The physics of low-dimensional semiconductors*, Cambridge University Press, New York (1998).

[40] A R Sugg and J P C Leburton, *IEEE J Quantum Electron.*, **27**, 224 (1991).

[41] A. K. Ghatak, K. Thyagarajan and M. R. Shenoy, *Journal of Lightwave Technology*, **LT-5**, 660 (1987).

[42] M. Grupen and K. Hess, *IEEE Journal of Quantum Electronics*, **34**, 1 (1998).

[43] Z-M Li, S. P. McAlister and C. M. Hurd, *Semicond. Sci. Technol.*, **5**, 408 (1990).

[44] P. N. Racec, U. Wulf and J. Kucera, *Solid-State Electronics*, **44**, 881 (2000).

[45] N. Ben Abdallah, *J. Stat. Phys.*, **90**, 627 (1998).

[46] P. Bordone, M. Pascoli, R. Brunetti, A. Bertoni and C. Jacoboni, *Phys. Rev. B*, **59**, 3060 (1999).

[47] L. F. Register and K. Hess, *Appl. Phys. Lett.*, **71**, 9 (1997).

[48] R. M. Kolbas, Y. C. Lo and J-H Lee, *IEEE J. Quantum Electron.*, **26**, 25 (1990).

[49] H. Schneider and K. Klitzing, *Phys. Rev. B*, **38**, 6160 (1988).

[50] D. Oriato, A B Walker and W N Wang, *VLSI Design, Proceedings of IWCE Glasgow in press*, (2000).

[51] F. Bernardini, V. Fiorentini, D. Vanderbilt, *Phys. Rev. B*, **56**, R10024 (1997).

[52] L. Görgens, O. Ambacher, M. Stutzmann, C. Miskys, F. Scholz and J. Off, *Appl. Phys. Lett.*, **76**, 577 (2000).

[53] R. Cingolani, A. Botchkarev, H. Tang, H. Morkoc, G. Traetta, G. Coli, M. Lomascolo, A. Di Carlo, F. Della Sala and P. Lugli, *Phys. Rev. B*, **61**, 2711 (2000).

[54] A. F. Wright, *J. Appl. Phys.*, **82**, 2833 (1997).

[55] C. Wetzel, S. Nitta, T. Takeuchi, S. Yamaguchi, H. Amano and I. Akasaki, *MRS Internet J Nitride Semicond. Res.*, **3**, 31 (1998).

[56] A. D. Bykhovski, B. Gelmont and M. Shur, *J. Appl. Phys.*, **74**, 6734 (1993).

[57] A. D. Bykhovski, V. V. Kaminski, M. S. Shur, Q. C. Chen and M A Khan, *Appl. Phys. Lett.*, **68**, 818 (1996).

[58] M. Schuster, P. O. Gervais, B. Jobst, W. Hösler, R. Averbeck, H. Riechert, A. Iberl and R. Stömmer, *J Phys D: Appl. Phys.*, **32**, A56 (1999).

[59] F. Bechstedt, U. Grossner, J. Furthmüller, *Phys. Rev. B*, **62**, 8003 (2000).

[60] T. Takeuchi, C. Wetzel, S. Yamaguchi, H. Sakai, H. Amano, I. Akasaki, Y. Kaneko, S. Nakagawa, Y. Yamaoka and N. Yamada, *Appl. Phys. Lett.*, **73**, 1691 (1998).

[61] S. I. Stepanov, W. N. Wang, B. S. Yavich, V. E. Bougrov, Y. T. Rebane and Y. G. Shreter, *MRS Internet J Nitride Semicond. Res.*, **6**, 1 (2001).

Chapter 3

GUNN Diode

The Gunn diode is a two terminal device which converts dc power into radiofrequency (rf) power with a domain of applications covering 10 to over 300 GHz. It achieves this using the negative differential resistance characteristics of bulk semiconductor such as GaAs or InP. Gunn diodes remain one of the most common high-frequency solid state oscillators at present and are considered as promising devices for millimetre and submillimetre-wave applications [1]. In addition market opportunities for Gunn diodes in millimetre-wave applications in intelligent cruise control/collision avoidance car radar and in local multipoint distribution systems (LMDS) are nowadays unprecedented.

The combined requirement of high power output (> 10's mW), frequency agility, wide band width, low output noise and wide temperature operation (-80 °C to 150 °C) places severe demands on the performances of the Gunn diodes. For that reason modelling is regarded as a key issue and fast simulation programs become more and more important.

In this chapter we will summarise the theory of Gunn diodes with special emphasis on the transferred electron phenomena. In section 3.1.1 we explain qualitatively how a charge instability forms in the presence of negative differential resistance (NDR). The time in which the charge build up is calculated in section 3.1.2. A simple model is presented in section 3.1.3. Section 3.1.4 deals with the dynamic of these charge instabilities called *domains*. Sections 3.1.5, 3.1.6 and 3.1.7 describe typical characteristics of domains: the equal-areas rule, the shape of domains

and their transit time. Section 3.1.8 introduce the concept of *modes*. The final section explains how NDR can be achieved in GaAs semiconductors.

We want to warn the reader of this chapter that the symbol for the electric field used in the text is $\mathcal{E}$ consistently with the chapter 2 and in order to distinguish it from the energy. On the other hand in the pictures for graphical reasons the electric field has been reported using the notation E.

3.1 Negative Differential Resistance and Charge Instability

In 1963, Gunn [2] discovered that coherent microwave output was generated when a dc electric field was applied across a randomly oriented, short, n-type sample of GaAs or InP that exceeded a critical threshold value of several thousand volts per centimetre. The frequency of oscillation was approximately equal to the reciprocal of the carrier transit time across the length of the sample. Later, Kroemer [3] pointed out that all the observed properties of the microwave oscillation were consistent with a theory of negative differential resistance (NDR), which had been independently proposed by Ridley and Watkins [4] and by Hilsum [5]. For a brief history and review of NDR, refer to [6].

3.1.1 Domain Formation

What happens when there is NDR? We will try to answer this question first in a qualitative way, then we will introduce a simple model to have a quantitative estimation of the variables involved.

A semiconductor exhibiting bulk negative differential resistivity is inherently unstable, because a random fluctuation of carrier density at any point in the semiconductor produces a momentary space charge that grows exponentially in time. Let us imagine a long line of identical electrons moving throughout the device. Consider the extreme situation where at some critical field E_t the electron velocity suddenly drops from v_t to v_s as in figure 3.1(a).

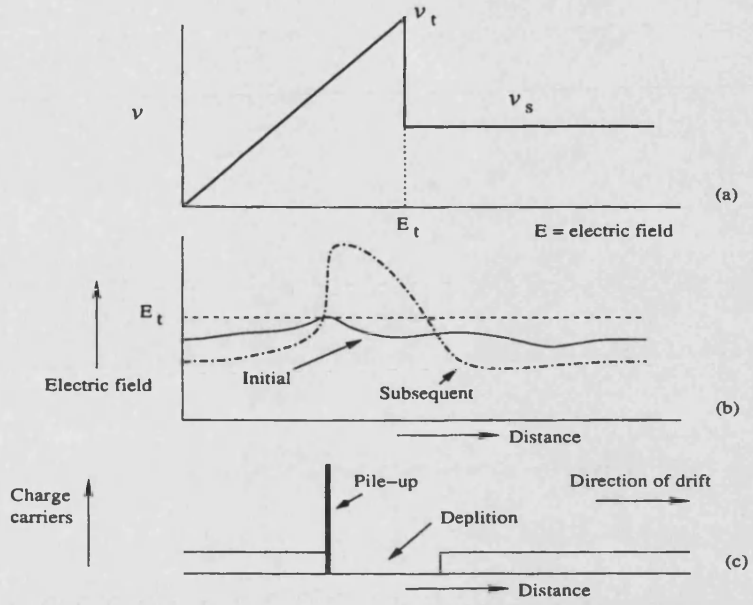

Figure 3.1: Domain formation

The electron that experiences this increased field consequently slows down. The electrons behind pile into it while the electrons in front move away from the pile-up leaving a region depleted of electrons. A local imperfection in the semiconductor where the field is slightly higher will thus nucleate an accumulation of charge following a depletion region as shown in figure 3.1(c). The change of the charge distribution from its normal state leads to a region of high electric field or a domain as it is called. This absorbs the applied potential across the device and the field outside the domain actually falls below the threshold value and so no more domains may form. This lowering of the field outside the domain leads to a lowering of the charge carrier velocity and so to a reduction in the device current. The domain travels with a velocity close to the saturated electron velocity v_t until it reaches the anode where it must disappear. The field in the sample then rises momentarily, and with it the current, until a new domain is formed and the process repeats.

3.1.2 Dielectric relaxation time

An important quantity in our work is the time scale on which charge can rearrange itself in a semiconductor. This time scale is given by the *dielectric relaxation* time. In an one-dimensional picture, the movement of an excess of charge Q above the

equilibrium n_0, is governed by charge continuity and Gauss' theorem. Combining these two laws yields

$$\frac{\mathrm{d}Q}{\mathrm{d}t} = -\omega_c Q, \qquad \omega = \frac{qn_0\mu}{\varepsilon}, \qquad \mu = \frac{\mathrm{d}v}{\mathrm{d}\mathcal{E}} \tag{3.1}$$

where q is the electron charge, ε the dielectric constant, v the carrier velocity and $\mathcal{E}$ the electric field. In a conductor or semiconductor the charge decays towards its equilibrium value in a time scale given by $\tau_c = 1/\omega_c$. τ_c is called the dielectric relaxation time and for GaAs is around 10^{-12} sec. If, however, the differential mobility μ is negative then the charge spontaneously grows away from the equilibrium condition. It is this that leads to domain formation. The dielectric relaxation sets the time scale for domain growth.

3.1.3 Simple model of domain formation

We consider now the way in which charge can redistribute itself in a semiconductor. The drift, diffusion and displacement currents for electrons combine to give the total density of current J in a one-dimensional analysis as

$$J = -qnv(\mathcal{E}) + q\frac{\partial(Dn)}{\partial z} + \varepsilon\frac{\partial\mathcal{E}}{\partial t} \tag{3.2}$$

Suppose then that the semiconductor has a donor density n_0 and a uniform field $\mathcal{E}_0$ except for a local non-uniform field region between the point z_1 and z_2. In terms of the uniform field, the total current density is

$$J = -qn_0v(\mathcal{E}_0) + \varepsilon\frac{\partial\mathcal{E}_0}{\partial t}. \tag{3.3}$$

Between the points z_1 and z_2 the changes in field and charge density are related by Gauss' theorem

$$\varepsilon \frac{\partial \mathcal{E}}{\partial z} = -q(n - n_0). \tag{3.4}$$

The total current is continuous so that equations 3.2 and 3.3 represent the same quantity and can be subtracted. Using equation 3.4 as well we obtain

$$\varepsilon \frac{\partial(\mathcal{E} - \mathcal{E}_0)}{\partial t} = -qn_0[v(\mathcal{E}_0) - v(\mathcal{E})] - q\frac{\partial(Dn)}{\partial z} - \varepsilon v(\mathcal{E})\frac{\partial \mathcal{E}}{\partial z} \tag{3.5}$$

Integrating over the non-uniform region from z_1 and z_2, noting that these points are just in the uniform field, we obtain a general equation given by Kurokawa [7]

$$\frac{\partial}{\partial t}\int(\mathcal{E} - \mathcal{E}_0)\mathrm{d}z = -\frac{qn_0}{\varepsilon}\int[v(\mathcal{E}_0) - v(\mathcal{E})]\mathrm{d}z. \tag{3.6}$$

When the field changes are small we can use a first order approximation for the velocity as a function of the field and substituting we get

$$\frac{\partial V_D}{\partial t} = -\omega_c V_D \tag{3.7}$$

where

$$V_D = \int(\mathcal{E} - \mathcal{E}_0)\mathrm{d}z \tag{3.8}$$

$$\omega_c = \frac{qn_0}{\varepsilon}\left(\frac{\mathrm{d}v}{\mathrm{d}\mathcal{E}}\right). \tag{3.9}$$

V_D is the excess voltage over and above the voltage $\mathcal{E}_0 L$ for the semiconductor bulk of length L. It is attributable solely to the disturbance in the sample. It can be seen that disturbances normally decay so there is no excess voltage. The time constant for the decay is the dielectric relaxation time. When, however, the differential mobility is negative the disturbance grows and this is what forms a domain. In figure 3.2 we present the simple idea of a *dipole domain*, called in this way because of the dipole of charge involved. In the limit of negligible diffusion it

has a sharp triangular shape with the accumulation layer occupying a negligible distance.

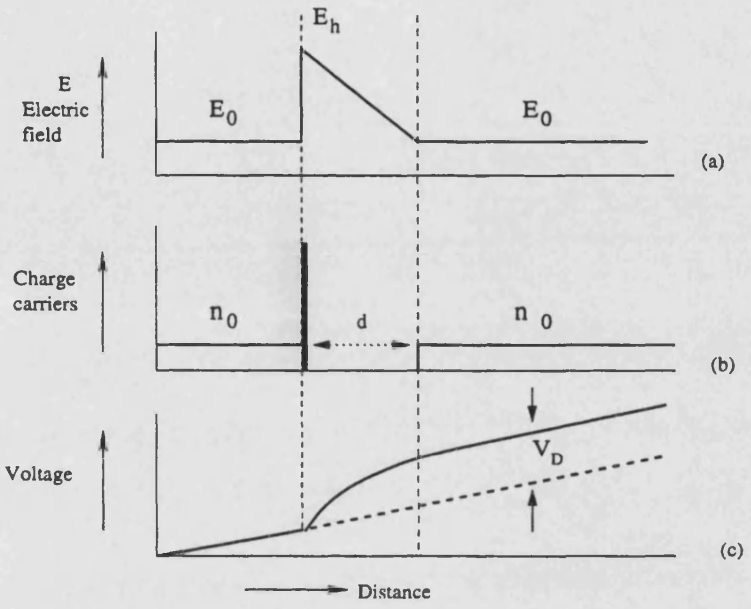

Figure 3.2: (a) Electric field, (b) charge density, (c) electrostatic potential for a dipole domain versus the device length.

3.1.4 Domain as a travelling wave

When considering difficult non-linear problems, it is often a good starting point to consider small changes from some equilibrium condition. This small signal approach is very useful when applied to drifting electron streams in semiconductors and shows us that, like any fluid flow, the drifting particles support waves. The use of a small signal study is described in chapter 2. We have shown that for spotting domain formation and motion, an approach involving small linear variation was necessary in the numerical treatment when the standard exponential approach failed.

We begin with the equations determining the motion of the disturbances along the stream of electrons in a semiconductor. The sign convention for a one-dimensional flow is represented in figure 3.3. The following relations must hold:

$$V = -\int \mathcal{E} \mathrm{d}z \tag{3.10}$$

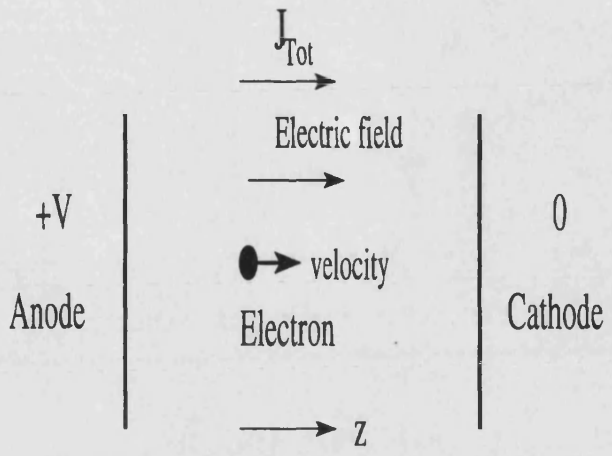

Figure 3.3: Sign convention

where V is the electrostatic potential across the device. The continuity equation can be written as

$$-q\frac{\partial n}{\partial t} + \frac{\partial J_{CD}}{\partial z} = 0 \tag{3.11}$$

J_{CD} being the particle current sum of convection and diffusion current

$$J_{CD} = -qnv(\mathcal{E}) + qD\frac{\partial n}{\partial z}. \tag{3.12}$$

The total current J relation is

$$J = J_{CD} + \varepsilon\frac{\partial \mathcal{E}}{\partial t}. \tag{3.13}$$

The total current flow is independent of position and is only a function of time in one dimension as it can be easily seen by showing that $\partial J/\partial z = 0$ directly from the previous equations. The diffusion coefficient has been taken to be a constant. It can be shown [8] that if energy transport is taken into account then extra wave solutions are found. However, these extra waves are often rapidly attenuated with distance and have little physical effect. A formal steady state solution to these equations when the doping density N_0 is uniform is given by

$$\mathcal{E}_0 = -V_0/L, \qquad J_0 = -qn_0v_0, \qquad n_0 = N_0 \qquad v_0 = v(\mathcal{E}_0) = -\mu_0\mathcal{E}_0. \quad (3.14)$$

On top of this steady solution some perturbations are superimposed: $\mathcal{E}_1$, n_1, etc. Products of terms involving the squares of perturbation (n_1v_1, etc.) are ignored to linearise the previous equations. The Gauss theorem becomes

$$\varepsilon\frac{\partial\mathcal{E}_1}{\partial z} = -qn_1 \quad (3.15)$$

while equations 3.11, 3.12, and 3.13 can be rewritten respectively

$$-q\frac{\partial n_1}{\partial t} + \frac{\partial J_1}{\partial z} = 0 \quad (3.16)$$

$$J_1 = -qn_0v_1 - qn_1v_0 + qD\frac{\partial n_1}{\partial z} \quad (3.17)$$

$$J = J_1 + \varepsilon\frac{\partial\mathcal{E}_1}{\partial t}. \quad (3.18)$$

Defining $\omega_c = (qn_0\mu_0/\varepsilon)$ and the Debye length $L_D = \sqrt{D/\omega_c}$, the previous equations can be rearranged to obtain the defining equation for the electric field in a normalised form:

$$\frac{J_1}{J_0} = \mu_r\left(\frac{\mathcal{E}_1}{\mathcal{E}_0}\right) + \frac{v_0}{\omega_c}\frac{\partial}{\partial z}\left(\frac{\mathcal{E}_1}{\mathcal{E}_0}\right) - L_D^2\frac{\partial^2}{\partial z^2}\left(\frac{\mathcal{E}_1}{\mathcal{E}_0}\right) + \frac{1}{\omega_c}\frac{\partial}{\partial t}\left(\frac{\mathcal{E}_1}{\mathcal{E}_0}\right) \quad (3.19)$$

where

$$\mu_r = \frac{\mathcal{E}_0}{v_0}\frac{\mathrm{d}v}{\mathrm{d}\mathcal{E}}. \quad (3.20)$$

The relative importance of diffusion and dielectric relaxation can be estimated from the relative size of the parameters L_D and v_0/ω_c. Generally, diffusion is less important in high resistivity materials while dielectric relaxation effects dominate.

In low resistivity materials diffusion can dominate. Equation 3.19 gives a partial differential equation for the electric field along the length of the semiconductor. We can seek wave-type solutions varying as $\exp j(\omega t - \beta z)$. With this solution in equation 3.19 one finds the values for β of the complementary solutions

$$\mu_r + j\frac{\omega}{\omega_c} - j\frac{\beta v_0}{\omega_c} + \beta^2 L_D^2 = 0. \tag{3.21}$$

The quadratic equation 3.21 can be solved approximately to give

$$\beta_f \sim \frac{\omega}{v_0} - j\mu_r\frac{\omega_c}{v_0} - j\frac{\omega_c}{v_0}\left(\frac{L_D\omega_c}{v_0}\right)^2\left(\frac{\omega}{\omega_c} - j\mu_r\right) \tag{3.22}$$

$$\beta_r \sim -\frac{\omega}{v_0} + j\mu_r\frac{\omega_c}{v_0} + j\frac{\omega_c}{v_0}\left(\frac{v_0}{L_D\omega_c}\right)^2 \tag{3.23}$$

where it has been assumed that $L_D\omega_c/v_0 << 1$.

If for the moment we ignore the imaginary part of the propagation constant (attenuation) and remember that ω/β and $\mathrm{d}\omega/\mathrm{d}\beta$ give the phase and group velocities respectively, then it is immediately seen the first wave (β_f) represents a forward wave and the second wave (β_r) travels in the reverse direction. Both waves have the same magnitude of phase and group velocity which are equal to the carrier drift velocity. Now if attenuation is considered, both waves are seen to be attenuated (for $\mu_r > 0$) as they travel along in their directions of propagation. Since it has been assumed that $L_D\omega_c/v_0 << 1$ and $\omega < \omega_c$ at practical frequencies and doping levels, it is clear that the reverse wave is very rapidly attenuated in a fraction of the wavelength. The attenuation of the forward wave on the other hand is low with a high resistivity (low ω_c) and/or with a low μ_r. Moreover a negative differential mobility with a large enough magnitude to swamp the attenuation caused by diffusion will enable the wave to grow rather than decay as it travels.

3.1.5 Equal areas rule

The previous analysis is completely general and says nothing about the shape of the disturbance, or domain. A more specific approach has to be used where diffusion mechanisms and the effects of a finite area of semiconductor have to be taken into account. In general, one finds that the mathematical complications become severe [8]. However an important result can be derived with not much effort. We will prove that the peak field in the domain and the external field can be related in the steady state by a graphical construction known as the *equal areas rule.*

Let us consider a uniform specimen of semiconductor with the usual one-dimensional analysis. The total current I through the specimen of area A is the sum of convection, diffusion and displacement currents. The density of current $J = I/A$ can be therefore expressed by equation 3.2. We now seek a solution of equation 3.2 that reflects the experimental observation that the field is a moving pattern; thus $\mathcal{E} = \mathcal{E}(z - ut)$. To solve for this situation we do a coordinate transformation to a moving system where the new coordinates are $z' = z - \int_0^t u(t)dt$ and $t' = t$. The derivatives in the old coordinates are expressed with the new coordinates as

$$\frac{\partial}{\partial t} = \frac{\partial}{\partial t'} - u\frac{\partial}{\partial z'} \qquad and \qquad \frac{\partial}{\partial z} = \frac{\partial}{\partial z'} \tag{3.24}$$

Equation 3.2 then becomes with the new coordinates

$$J = -qnv(\mathcal{E}) + qD\frac{\partial n}{\partial z} - u\varepsilon\frac{\partial \mathcal{E}}{\partial z} + \varepsilon\frac{\partial \mathcal{E}}{\partial t}. \tag{3.25}$$

where the primes on the new coordinates have been dropped for convenience. Gauss theorem in the new moving frame is unaltered in form:

$$\varepsilon\frac{\partial \mathcal{E}}{\partial z} = -q(n - n_0) \tag{3.26}$$

Equation 3.25 can be simplified by considering the variable to be the field $\mathcal{E}$ instead of the distance z. This change comes from equation 3.26, $\partial/\partial z = -[q(n - n_0)/\varepsilon]\partial/\partial\mathcal{E}$. Substituting in equation 3.25 and integrating from the external field $\mathcal{E}_0$ to the peak field in the domain $\mathcal{E}_h$ we obtain the integral relation

$$\frac{qn_0D}{\varepsilon}\int_{n(\mathcal{E}_0)}^{n(\mathcal{E}_h)}\left(1-\frac{n_0}{n}\right)\frac{\mathrm{d}n}{n_0} = \int_{\mathcal{E}_0}^{\mathcal{E}_h}[v(\mathcal{E})-\sqcap]\mathrm{d}\mathcal{E}+\int_{\mathcal{E}_\prime}^{\mathcal{E}_\wr}\frac{(\mathcal{J})-\varepsilon\frac{\partial\mathcal{E}}{\partial\sqcup}+\amalg\backslash_\prime\sqcap}{\amalg\backslash}\mathrm{d}\mathcal{E} \quad (3.27)$$

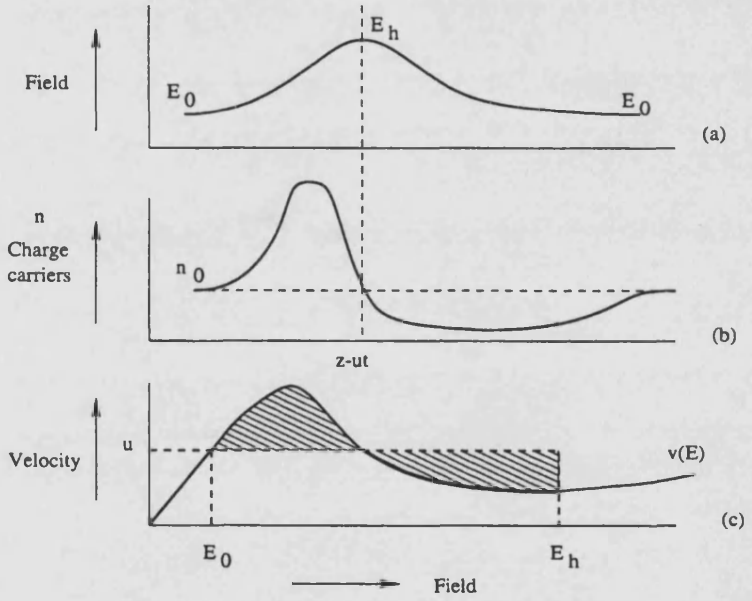

Figure 3.4: Equal areas rule: (a) electric field, (b) carrier density, (c) carrier velocity.

Now we consider only the steady state case of equation 3.27 ($\partial/\partial t = 0$) in the moving frame of reference. It is clear that the maximum field $\mathcal{E}_h$ occurs at the point where $n = n_0$ as the accumulation of charge changes over to a depletion as shown in figure 3.4(a)(b). Thus the lefthand integral in equation 3.27 vanishes for either of two possible paths of integration of $\mathcal{E}$ from $\mathcal{E}_0$ to $\mathcal{E}_h$: over the accumulation or alternatively over the depletion region. The first integral on the righthand side of equation 3.27 is a function only of $\mathcal{E}$ and is therefore independent of which path is taken for the integration. The last integral has a constant on top but a variable on the bottom. It can be seen in figure 3.4(b) that the variable n is always greater than n_0 along one path and always less than n_0 along the other. It follows that the only way that this integral can be independent of path is for the numerator to vanish: $[J + qn_0u] = 0$. One concludes that the external electron velocity equals the domain velocity since

$$J = -qn_0 v(\mathcal{E}_0) = -qn_0 u \qquad so \qquad u = v(\mathcal{E}_0). \tag{3.28}$$

This leave the integral

$$\int_{\mathcal{E}_0}^{\mathcal{E}_h} [v(\mathcal{E}) - u] \mathrm{d}\mathcal{E} = 0 \tag{3.29}$$

relating the peak domain field to the external field. The geometrical interpretation of this rule as two equal areas on the $v(\mathcal{E})$ characteristic is given in 3.4(c) and was first given by Butcher [9].

3.1.6 Domain patterns

In principle there is no reason why the accumulation region and the depletion region of a domain should not be separated by a distance in which the carrier density is uniform at the donor density value. The field in this region would be uniform, leading to a *flat-topped* profile for the distance plot as in figure 3.5(a).

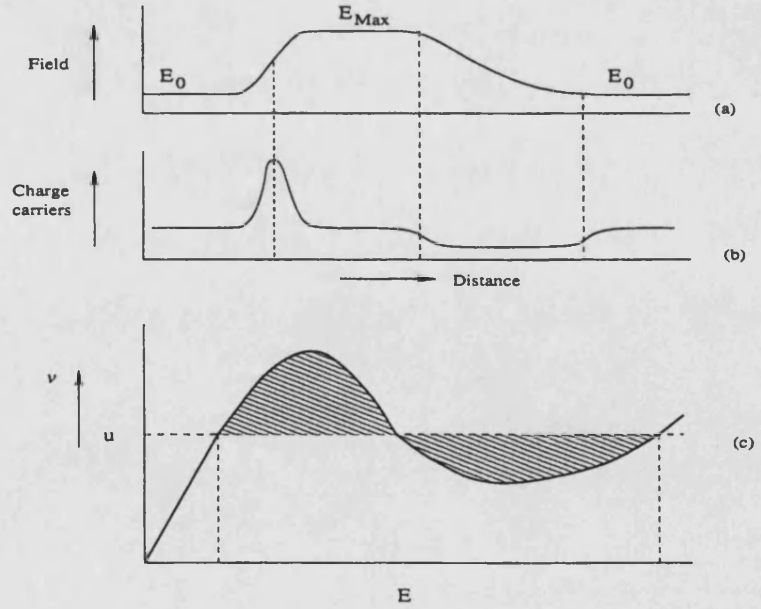

Figure 3.5: (a) Field and (b) charge patterns for flat-topped domain; (c) field relations for stable flat-topped domain.

It can be shown [8] that this type of domain is unstable except when the high

field velocity on the $v(\mathcal{E})$ characteristic is equal to the low field velocity. The flat-topped domain can then occur in a stable form only in material with a $v(\mathcal{E})$ characteristic of the type shown in figure 3.5(c), where the velocity rises again at high enough fields. Such stable flat-topped domains have not been observed in GaAs for fields up to 120 kV/cm. In fact calculated velocity-field characteristics show that the electron velocity is indeed saturated up to this value of field.

When considering domain formation it has been assumed that there is inevitably some non-uniformity that leads to a dipole type of domain. In computer simulations it is usual to put this non-uniformity into the calculation by placing a notch of lower donor density near the cathode. If this non-uniformity is left out, calculations show that, after the field exceeds threshold, an accumulation of charge is created near the cathode and moves through the sample. The depletion that is required to ensure the overall electrical neutrality of the sample occurs instead at the anode contact. The charge and field profile for this type of domain, called *accumulation domain* are shown in figure 3.6(a)(b). Since the high field is in a negative mobility region any small depletion will grow into a large depletion. On the other hand if the high field is in a positive mobility region as in figure 3.5(c), the accumulation domain is stable.

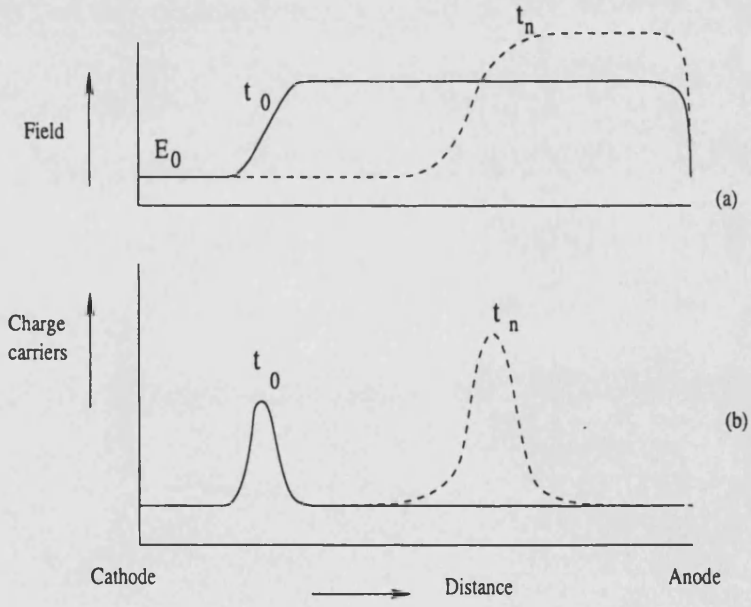

Figure 3.6: (a) Electric field E and (b) charge patterns for accumulation domain.

3.1.7 Doping length and time scale

For a given doping density there will be a minimum device length that will support a domain due to the finite time required for the domain growth. If this time is longer than the time required for the domain to traverse the sample, then domain formation will not occur. The domain can be initiated by the device voltage exceeding the threshold value and in general it forms at the cathode. The time scale of the growth of the domain is governed by the dielectric relaxation time, $\tau_c = 1/\omega_c$ and in GaAs an approximated value of τ_c is given by the ratio $10^3/N_D$ where N_D is the donor density in cm^{-3} [8]. To talk sensibly about a domain being present it is clearly necessary for the domain to be able to grow significantly during the transit time $\tau_t = L/v_t$, where L is the length of the specimen and v_t is the velocity of the domain. In general calculations show [8] that a useful order of magnitude for a maximum growth time is around $50\tau_c$ so that this limitation yield the relation $\tau_t = L/v_t < 50\tau_c$. In terms of N_D we obtain a famous rule valid for GaAs material [8]:

$$N_D L > 10^{12}\mathrm{cm}^{-2}. \tag{3.30}$$

It has been shown that a device with subcritical $N_D L$ product is usually stable [10]. However, if the doping density is too high, the current (and hence the temperature) in the semiconductor becomes too elevated and the life expectancy of the device is reduced. In practice, the product of doping density and device length is maintained between the following limits:

$$10^{12}\mathrm{cm}^{-2} < N_D L < 2 \cdot 10^{12}\mathrm{cm}^{-2}. \tag{3.31}$$

3.1.8 Modes

The domain can be extinguished in two ways once it has been formed: it can either travel right through the device and run into the anode contact, *transit time mode*, or the device voltage can fall to a low enough value so that the domain discharges during transit before it reaches the anode, *quenched modes*.

In addition, we can classify the transit time mode in three other modes:

- The proper transit time mode is the basic diode oscillation mode and is independent of the external circuit. A peak in the current is obtained when a domain is quenched at the anode, after which another is nucleated near the cathode. The frequency is determined by the domain transit time τ_t: $f_t = 1/\tau_t$. However the transit time frequency depends sensitively on the operating voltage and temperature, therefore this mode has poor frequency stability and efficiency.

- The resonant transit mode is just a particular case of the previous mode. In this mode it is assumed that the domain transit time is precisely equal to the period of the resonant circuit coupled with the device. The frequency is still given by the domain transit time as in the proper transit time mode, but stability is highly increased by using the resonant circuit.

- The delayed domain mode sets in when the total electric field across the device drops below the threshold value, E_h, during part of the r.f. cycle of the external circuit, such that nucleation of a new domain is delayed. As soon as the field rises above E_h, a domain nucleates at the cathode and travels across the device. As the field swings below E_h, the domain arrives at the anode and dumps its charge. A new domain cannot form at the cathode until the field rises above E_h again. This delay time τ_d between extinction and creation of domains modifies the operating frequency, $f = 1/(\tau_t + \tau_d)$. The delay time is a function of the r.f. voltage, which is determined by the external circuit. It follows that the operating frequency is always below the transit time frequency, it can be made very stable and it can be tuned on a broader bandwidth.

3.2 Transferred Electron Devices

The mechanism responsible for the Negative Differential Resistance (NDR) in several semiconductors was found to be the transferred electron effect. The transferred electron effect is the transfer of conduction electrons from the high-mobility

low-energy valley to low-mobility higher-energy satellite valleys [11]. The main characteristics are:

- there is a well defined threshold field for the onset of NDR;
- the threshold field increases with the lattice temperature;
- the NDR can be destroyed by having the lattice temperature too high or the energy difference between valleys too small;
- the lattice temperature must be low enough that, in the absence of a bias electric field, most electrons are in the lower conduction-band minimum;
- in the lower conduction-band minimum the electrons must have high mobility, small effective mass and low density of states whereas in the upper satellite valleys the electrons must have low mobility, large effective mass and high density of states;
- the energy separation between the two valleys must be smaller than the semiconductor bandgap so that avalanche breakdown does not set in before electrons are transferred into upper valleys.

Of the semiconductors satisfying these conditions, n-type GaAs and n-type InP are the most widely studied and used. It is expected that when we increase the external electric field E, the carrier velocity will at first increase in accordance with the relation $v = \mu E$ where μ is the carrier mobility. For instance in GaAs, the electrons reach a peak velocity of $v_{pk} = 2.2{\cdot}10^7$ cm/s for an electric field $E \simeq$ 3 kV/cm before intervalley transfer occurs from the Γ-valley to the $L(X)$-valley and, then, decrease to the saturated values of $v_s = 1{\cdot}10^7$ cm/s. Figure 3.7 shows the electron velocity versus the electric field as obtained from a Monte Carlo simulation [12].

In alloy semiconductors, such as $Al_xGa_{1-x}As$, the electrons see potential fluctuations due to the compositional disorder (alloy scattering). This effect together with a slightly different band structure causes a deviation of the velocity-field curve from the starting material, eg GaAs. The saturated electron velocity in the central valley (i.e. the peak velocity v_{pk}) decreases drastically with x, but velocities in the X and L valleys, that are responsible for v_s, increase slightly. As result the absolute value of the NDR decreases.

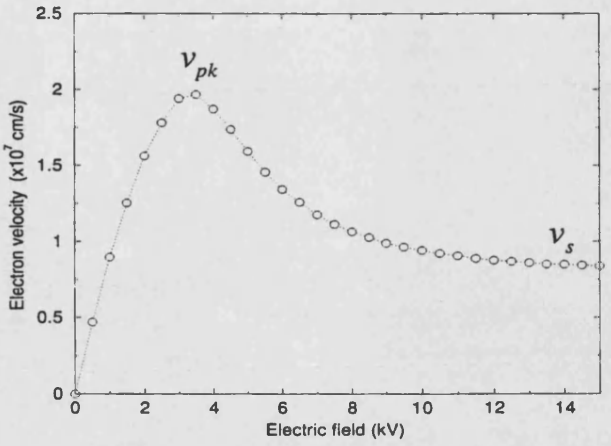

Figure 3.7: Monte Carlo Simulation

3.3 Conclusion

Even if many years have elapsed since that 1963 when the Gunn effect was discovered, the physics of transferred electron devices has been the object of constant attention. It has been possible to improve more and more the performances of commercial Gunn diodes thanks to the deep knowledge of the theory and to analytical calculations. Here we have presented what we consider the basic principles of charge instability in semiconductors by which optimum conditions for a simple Gunn diode can be deduced. However, the possibility of growing multi-heterostructure devices has brought a new degree of complexity. The equations are no longer analytically solvable and a computer simulation needs to be used as we will see in the next chapter. Nevertheless, the theory of this chapter remains the foundation of every Gunn-effect based device, where all the future developments start from.

References

[1] G. B. Tait and C.M. Krowne, *IEEE Transactions on Electron Devices*, **35**, 223 (1988).

[2] J. B. Gunn, *Solid State Commun.*, **1**, 88 (1963).

[3] H. Kroemer, *Proceeding IEEE*, **52**, 1736 (1964).

[4] B. K. Ridley and T.B. Watkins, *Proc. Physics Soc. Lond.*, **78**, 293 (1961).

[5] C. Hilsum, *Proc. IRE*, **50**, 185 (1962).

[6] N. Balkan et al., *Negative Differential Resistance and Instabilities in 2-D Semiconductors*, Plenum Press, New York (1993).

[7] K. Kurokawa, *Bell Syst. Tech. J.*, **46**, 2235 (1967).

[8] J. E. Carroll, *" Hot electron microwave generators,"* Edward Arnold, London, (1970).

[9] P. N. Butcher, *Phys. Letters*, **19**, 546 (1965).

[10] H. Thim and J. Turner, *"Handbook on Semiconductors"*, Edit. Chilsum (1993).

[11] S. M. Sze, *Physics of semiconductor Devices*, Wiley & Sons, New York Second Edition, (1981).

[12] S. J. Woods, *Simulation of Photoactivated Bipolar Devices*, PhD thesis, University of East Anglia, 1998.

Chapter 4

GUNN Diode Results

4.1 Introduction

Gunn diodes remain one of the most common high-frequency solid state oscillators at present and are considered as promising devices for millimetre-wave applications [1]. Extending the frequency range of a transit mode Gunn oscillator to over 100 GHz involves decreasing the Gunn diode active layer thickness (see section 3.1.8 in chapter 3). In this situation conventional $n^+ - n - n^+$ diodes (see next section) operate with reduced efficiency because part of the transit region (named *dead zone*) is used to accelerate and transfer electrons in the upper valleys *dead zone* shortening the effective transit region length. The dead zone length, for high frequency devices becomes comparable with the active layer thickness. The efficiency of submicrometre Gunn diodes operating in the transit mode may be increased by decreasing the *dead zone* by rapid electron acceleration to the intervalley transfer energy near the cathode. This rapid acceleration can be achieved by using a non-uniformly doped diode structure [2][3] or by hot electron injection from the cathode into the transit region [4][5]. In this case the *dead zone* length decreases to some hundreds of angstroms and the transit domain can form completely. New phenomena, such as hot-electrons effects, complicate the calculation of Gunn diode performance and prediction of ultimate device characteristics. Under these conditions numerical simulation of carrier transport is the only adequate method for submicrometre Gunn diode analysis. We have developed a one-dimensional time-dependent drift-diffusion model as described in

chapter 2 and we have used it in order to improve existent Gunn diode structures and to propose new solutions. The chapter is divided in two main sections: the simulation of a conventional Gunn diode and the simulation of the Hot Electron Injector Gunn Diode (HEIGD).

4.2 Conventional Gunn Diode

A conventional Gunn diode generally consists of three layers; a relatively low doped transit region sandwiched between two highly doped contact regions, forming a $n^+ - n - n^+$ structure as in figure 4.1.

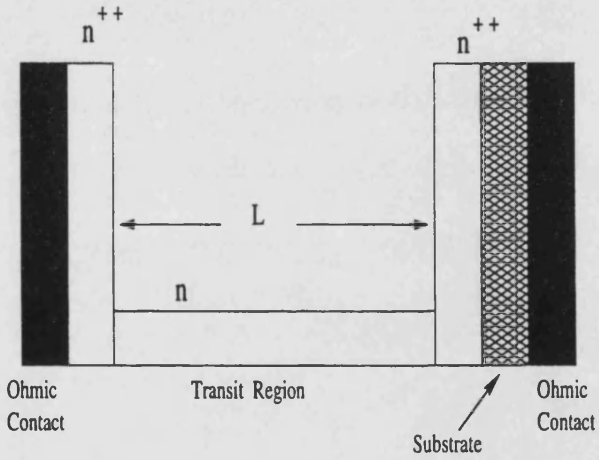

Figure 4.1: Schematic structure for a conventional Gunn diode.

The frequency of the Gunn diode is determined primarily by the transit region length L. However, a portion of this region is used to accelerate the electrons from the cathode until they have sufficient energy to enter the low mobility state and does not support domain formation. This *dead zone* may be as much as 0.25 μm in a transit region length of 1.5 μm (millimetric diode) and, because it acts as a parasitic resistance, it results in reduced efficiency.

The approximate behaviour of the electric field and carrier densities in a Gunn diode have been derived analytically in chapter 3. There, we observed that domains form near the cathode helped by charge inhomogeneities due to impurities or random fluctuation of the doping. In order to take account of this charge unbalance, we have modified the conventional $n^+ - n - n^+$ structure including a 0.25 μm layer of GaAs near the cathode. This is a necessary step that has been

reported also by other authors [6][7] and it outlines how computer models are limited in describing physical phenomena that have a random nature (at least in their early stage). Without this doped layer, simulations show no domain Gunn diode oscillations.

Here, for the purpose of demonstrating the validity and usefulness of our drift-diffusion method, the simulation has been carried out on a 3 μm long n-GaAs sample at temperature T_0=300 K with two different doping inhomogeneities at the cathode.

4.2.1 Device A: Accumulation-layer mode

Figure 4.2 shows the schematic structure of our simulated device (called Device A). A 0.25 μm-long doping notch of $7{\cdot}10^{15}$ cm^{-3} at the cathode acts as a source of carriers. An accumulation-layer is expected to form and propagate along the 2.75 μm length of a GaAs layer doped to a lower donor density. The device has been biased in its negative differential resistance region by a voltage of 2.1 V. We have used our time-dependent drift diffusion model with linear discretisation; the field-velocity curve was obtained with the MonteCarlo solver and it is reported in figure 3.7. Diffusion coefficients have been kept constants and are reported in appendix B with all the other parameters used.

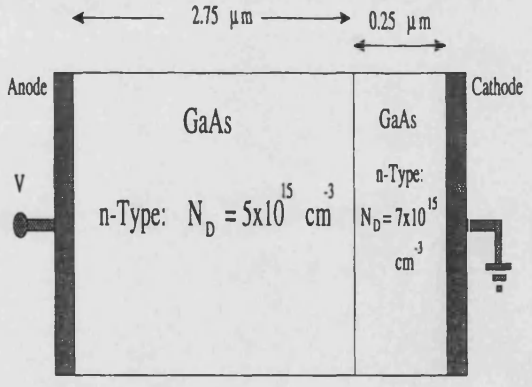

Figure 4.2: Device A: accumulation-layer mode

Figure 4.3 shows the electron density of the device of figure 4.2. An accumulation layer propagates from cathode to anode, where it becomes stationary. Figures 4.4 and 4.5(Device A labelled curve) show, respectively, the electric field and current as a function of the layer position and of the time. A small accumulation layer is

formed from the layer near the cathode doped at $7{\cdot}10^{15}$ cm^{-3}. Since the diode is biased in its negative differential mobility region, this small accumulation layer grows as it propagates, with electron velocity upstream of the layer greater than the velocities downstream. The stability of this layer at the anode is caused by a back diffusion of space charge counterbalancing the outflow of charge. The resulting electric field distribution is a stationary high field at the anode, with the field below threshold (roughly 3.5 $kVcm^{-1}$) in the remainder of the device as shown in figure 4.4. The maximum field strength observed is approximately 40 $kVcm^{-1}$ at the anode boundary.

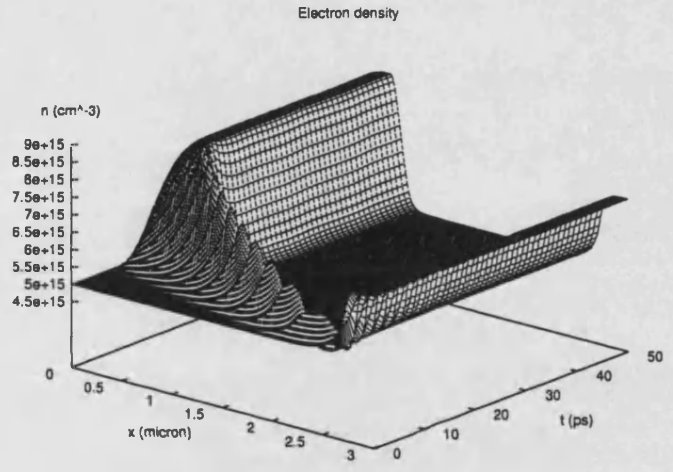

Figure 4.3: Electron density in device A as function of the position $x(\mu m)$ and time (ps). The anode is positioned at $x{=}0$.

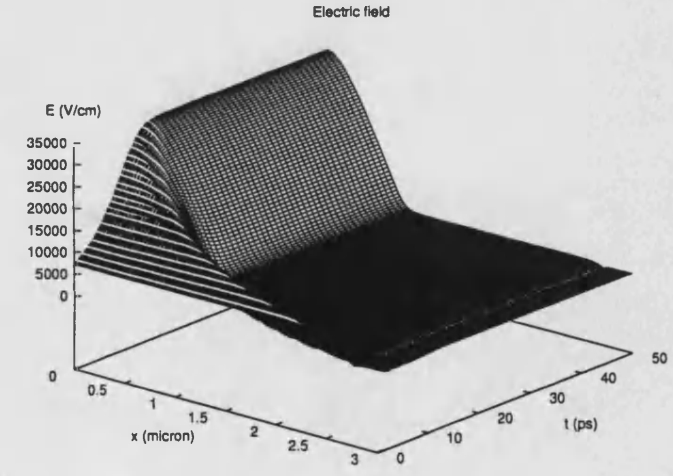

Figure 4.4: Electric Field in device A as function of the position x (μm) and time (ps)

A steady current value is achieved after approximately 25 ps (figure 4.5). Stable

current switching properties of over-critically doped GaAs diodes were predicted by early computer simulation using a constant diffusivity [8][9] and were verified by experiments [19].

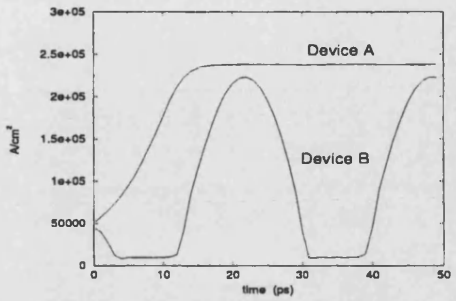

Figure 4.5: Current densities of device A and device B as function of time (ps)

The current calculation has been carried out adding the drift and diffusion components of the currents in a region of the device closed to the anode. This calculation neglects the effects of moving charges within the device.

4.2.2 Device B: dipole-layer mode

We have simulated a second device (Device B) as shown in figure 4.6. The layer near the cathode has a doping level of $3.5 \cdot 10^{15}$ cm^{-3}, lower than the doping of the 2.75 μm GaAs layer. All the others parameters remain the same as in the previous simulation.

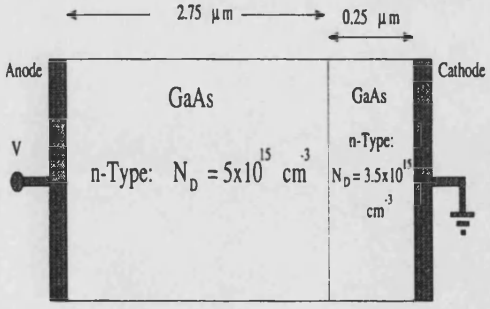

Figure 4.6: Device B: diode-layer mode

Now, the doping inhomogeneity at the cathode has the right sign and sufficient magnitude to lift the electric field up (see figure 4.8) and generate charge imbalances. These charge imbalances are dipole domains, which are characterised by short regions of high electric field as described in chapter 3. These dipole domains

grow to a stable size and drift toward the anode as shown by the electron density plotted in figure 4.7.

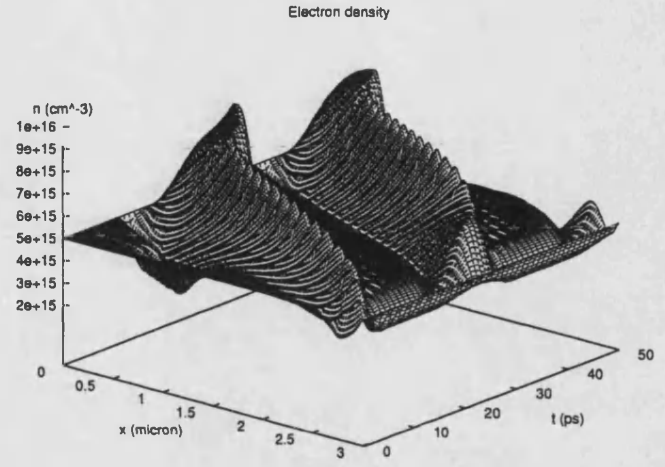

Figure 4.7: Electron density as function of the position x (μm) and time (ps)

As soon as the depletion zone of the dipole hits the anode, the electric field starts to rise above threshold back at the doping inhomogeneity, causing a renucleation of a small dipole domain. The excess charge at the anode quickly depletes as the new dipole grows. A cyclical formation of dipole domains results in current oscillations (pulse) at the dipole domain transit frequency. These well known RF short circuit Gunn effect oscillations are clearly illustrated in figure 4.7 and 4.8, showing, respectively, the electron density, n, and the electric field, E. The applied voltage is, again, 2.1 V.

The total terminal current pulses (see figure 4.5) have an oscillation of approximately 28 ps. This result is in good agreement with the basic theory [10] and previous Monte Carlo and Drift-Diffusion simulations [11].

4.3 Hot Electron Injector Gunn Diode

To maximise the power and efficiency for a particular length of the transit region of a $n^+ - n^- - n^+$ diode the accumulation layer should form as close to the cathode as possible. By choosing an appropriate transit region length and doping it is possible to maximise the power (at some bias) for a desired frequency. The

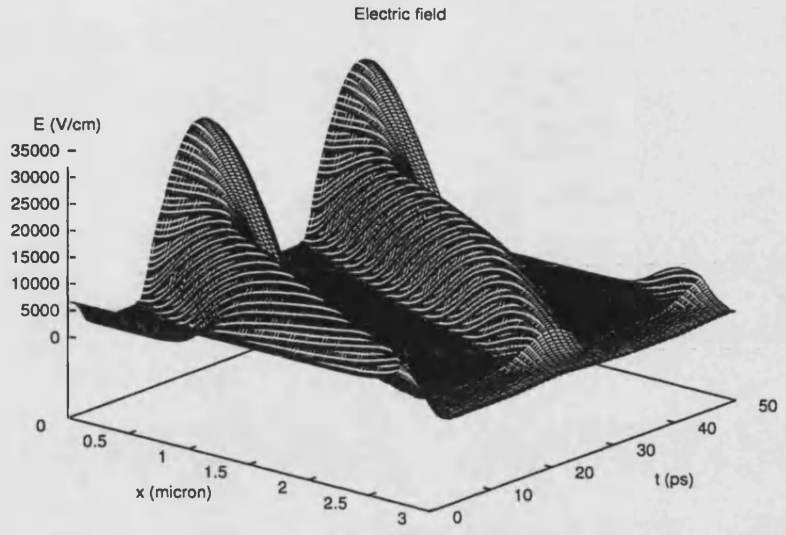

Figure 4.8: Electric field as function of the position x (μm) and time (ps)

problem for real applications is that the applied bias will vary, perhaps by as much as 1 V in ~5 V. In terms of the physics, when the voltage is reduced from the peak-power voltage, the point at which the accumulation layer forms moves away from the cathode thus shortening the effective transit length. This, in turn, increases the frequency of the oscillator. In reality, the diode is located in a high Q resonant cavity; the effect of increasing the diode frequency is that it is no longer possible to couple power out from the diode to the waveguide, and so coherent microwave power from the oscillator stops. It is possible to reduce the voltage dependence in $n^+ - n^- - n^+$ diodes, but only at the expense of a greatly reduced maximum power.

The solution is to design a Gunn diode where the critical field point is fixed at the cathode. To achieve this an injecting structure should be placed at the cathode. The purpose of the injector is to tailor the electric field, current and charge distribution at the cathode end of the transit region. The idea is to design a structure so that the accumulation layer starts at the cathode as soon as the electric field is high enough to sustain accumulation formation and propagation. This will produce a device with an effective transit length which is independent of the applied bias above some threshold bias.

There are several possible semiconductor structures which may be used as injectors. These include graded composition AlGaAs layers [12][13][14], planar doped

barriers [15][3], camel diodes [2], Schottky barriers [4] and tunnel injectors [21]. With the exception of tunnel injectors, all the others may be characterised as a triangular barrier placed at the cathode between the heavily-doped contact layer and the transit region. The ideal shape for the potential barrier is a relatively long arm nearest the contact layer with an abrupt interface at the transit region. Planar doped barriers and camel diodes have non-abrupt interfaces because of depletion effects. It is notoriously difficult to engineer Schottky barrier heights. Tunnel injectors were discounted because of the problems of simultaneously attaining good injection energies and the necessary current densities. It has therefore been concluded that graded AlGaAs layers would be the best structure available for study.

4.3.1 Graded-Gap Injector

The simplest design of a GaAs/AlGaAs graded-gap Gunn diode is to place an undoped linearly-graded $Al_xGa_{1-x}As$ layer, where x increases linearly, for instance from x=0 to x=0.3 as shown in figure 4.9.

In the graded-gap Gunn diode electrons are injected into the transit region with an energy greater or equal to the $\Gamma - L$ intervalley separation so that stable domains will form very near to the cathode and move across the transit region as soon as the field is high enough to sustain accumulation and propagation. The dead zone is effectively eliminated and the transit length is fixed and independent of bias.

For a GaAs/AlGaAs graded-gap Gunn diode the injection energy is equal to the difference in the conduction band energy between GaAs and the peak AlGaAs composition. This is typically in the range 200 to 300 meV depending on the fraction of Al present. This value is close to the $\Gamma - L$ valley energy separation for GaAs.

We have simulated a device with a transit region of 1.8 μm with a n-doping of 10^{16} cm^{-3}. The two contact regions are n-doped 10^{18} cm^{-3} and 0.5 μm length. The graded-gap injector has been considered slightly n-doped 10^{14} cm^{-3} with a length of 0.2 μm. Figure 4.10 shows the electron density and electric field both at equilibrium and under applied bias conditions. The electric field produced

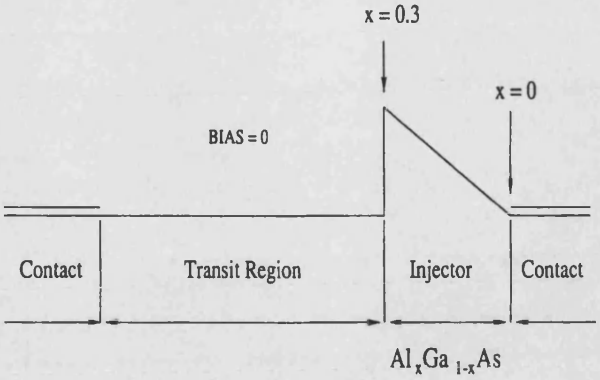

Figure 4.9: Graded-Gap Gunn Diode: schematic device structure.

when forward biased is rather high, $\sim 4 \cdot 10^4$ V/cm. This field causes depletion in the transit region over a length of about 500 nm. From the studies of hot electron transistors, it is known that a typical mean-free-path (mfp) in lightly doped GaAs is ~200 nm, falling quickly as the doping is increased. For this reason the hot electrons will have lost most of their energy by the time they have traversed the injector and the depletion layer. In the depletion region the electric field presents a positive gradient, ie the mobility decreases because of NDR effect. This means that electrons, moving in the opposite direction of the field, are subjected to an increasing mobility as they get into the transit region, preventing any accumulation. In the remaining part of the transit region the electric field is mainly below the threshold (dotted line in figure 4.10).

It is worth underlining the behaviour of electrons in the AlGaAs injector. This lightly doped region has a high density of electrons at the contact with the transit region. Such accumulation is not consistent with the presence of a barrier in the band structure. Further investigations, using Monte Carlo simulation, showed that the electric field in the AlGaAs injector is high enough to transfer electrons from the Γ band to the X band. Here electrons, having a lower mobility, accumulate as shown in figure 4.15 (dotted line). Figure 4.16 shows the percentage of electron population in the three bands, as confirmation at the heterojunction electrons are largely in the X band. The comparison between figure 4.10 and figure 4.15 shows good agreement between the drift-diffusion and the Monte Carlo results.

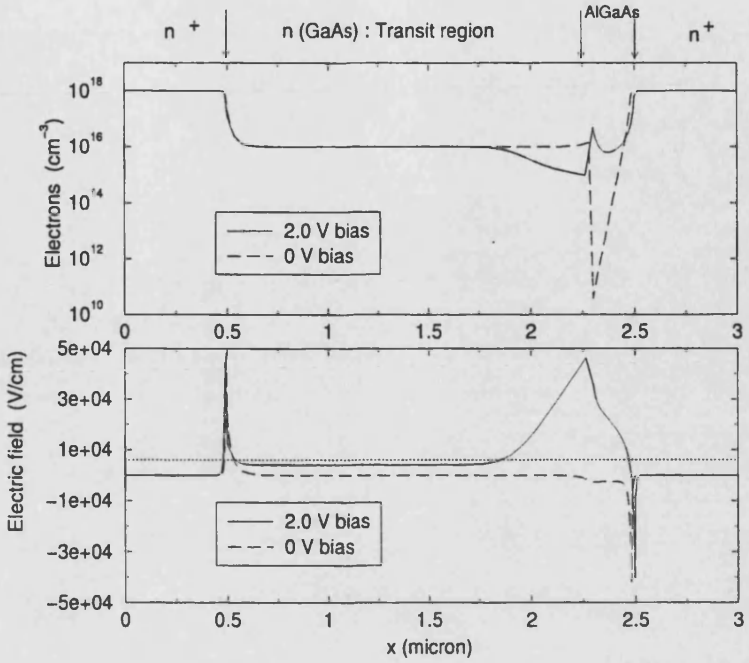

Figure 4.10: Graded-Gap Gunn Diode: electron density (top panel) and electric field (bottom panel) versus the position x (μm).

4.3.2 Graded-Gap Injector with n^+ spike

Both the reduction of the thickness of the depletion layer in the transit region and the tailoring of the curvature of the electric field is achieved by placing a highly doped n^+ layer between the abrupt interface of the graded-gap and the transit region as shown in figure 4.11.

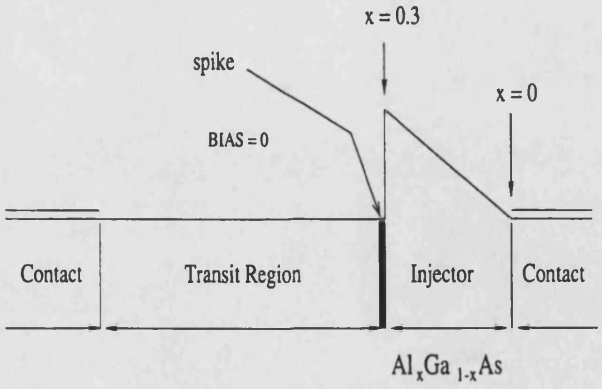

Figure 4.11: Graded-Gap Gunn Diode with n^+ spike: schematic device structure.

This acts as a non-equilibrium connector or short wire between the injector and the transit region. Being so thin, much less than the shortest mean free path, most hot electrons will pass through without any loss of energy. The correct

doping and thickness will create the appropriate electric field at the n^+ spike-transit region interface to sustain domain formation and propagation.

The device structure is the same as the one used for the simulation of figure 4.9 except for a spike made by a 200 nm GaAs layer with a doping of $1.5{\cdot}10^{17}$ cm^{-3}. The transit region is 1.8 μm length with a uniform doping of 10^{16} cm^{-3}. The grade-gap launcher is an $Al_xGa_{1-x}As$ layer with x=0.3 and x=0 at the interface with the spike and with the contact region respectively. The latter is 0.5 μm in length and 10^{18} cm^{-3} n-doped. The simulation of this structure is shown in figure 4.12. Both electron density and electric field profiles are plotted at zero bias (dotted line) and with an applied voltage of 2 V (solid line). A comparison with the simulation of figure 4.10 (without spike) reveals that the spike does not modify dramatically the characteristics of the zero-biased device. On the other hand the field and electron density are changed considerably under bias: in the transit region, for example, the depletion region has been reduced and the electric field is constant above the threshold (see figure 4.12). These are exactly the features required for domain propagation.

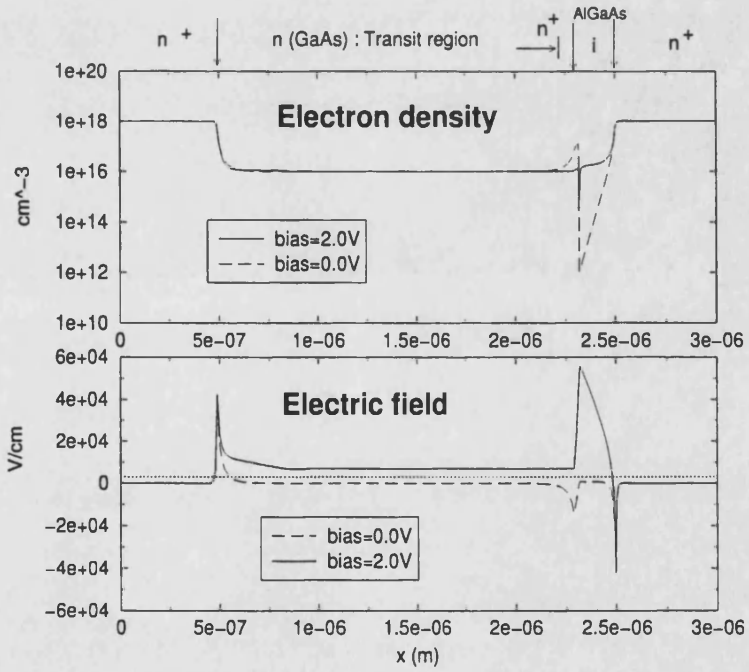

Figure 4.12: Graded-Gap Gunn Diode with n^+ spike: electron density (top panel) and electric field (bottom panel) versus the position x in the device.

To understand better the role of the spike further simulations have been carried out on the previous structure. First we have kept the applied voltage constant at a value of 2 V and varied the doping concentration of the spike. The results

are shown in figure 4.13. We can see that for low doping (solid line) the field has the typical triangular shape of the case without the spike. This means that there is still a depletion region in the transit region. A critical value is reached around a doping level of 10^{17} cm^{-3} where a slight change in doping results in a large change of the field. An optimum value is $1.5{\cdot}10^{17}$ cm^{-3} at which the field is constant and above threshold in all the transit region. Increasing the doping over the optimum value reduces the field in the transit region and increases it in the launcher. Furthermore an accumulation of electrons is formed near the anode which decreases the effective length of the transit region.

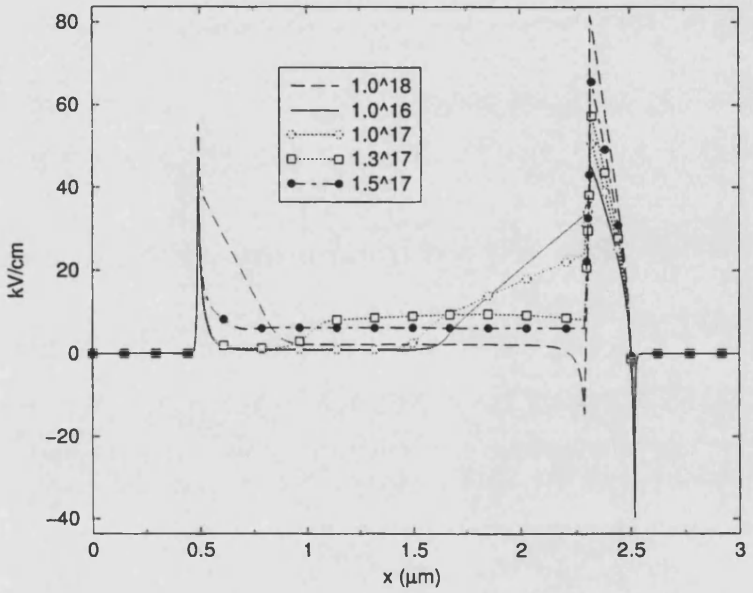

Figure 4.13: Electric field for different n^+ layer doping conditions versus the position x

The negative charge from electrons injected in the intrinsic AlGaAs region is, in the absence of a n^+ layer, balanced by a depletion region inside the transit region, i.e. by the fixed positive charge of ionised donors. The field does not increase uniformly across the transit region but has a triangular shape. As a consequence of the positive gradient of the field, electrons would accelerate, making the formation of an accumulation layer impossible. When we insert the thin high-doped spike between an intrinsic region (the launcher) and a low-doped transit region we expect a diffusion of electrons from the high density to the low one. Hence an unbalanced positive charge made of ionised donors is left in the spike. Therefore under bias conditions this positive charge will compensate the negative one injected from the cathode into the launcher and no depletion region will form in the transit region. As we expect and as we can see from figure 4.13 a low-doped spike does not bring sufficient fixed charge to avoid the formation of a depletion

region in the transit region. As soon as the doping is high enough, equilibrium of charge is reached and the field becomes flat and above threshold in the transit region.

A further step in our analysis has been to test the structure under different bias conditions. The doping in the spike has been kept at an optimum value of $1.5 \cdot 10^{17}$ cm^{-3}. Figure 4.14 shows the different profiles of the electric field. What we expect is that the charge inside the launcher increases as the applied bias increases due to injected electrons. For low bias the negative charge in the launcher is smaller than the fixed positive charge of the spike. Increasing the voltage causes the injected electrons to reach an equilibrium value that depends on the doping density of the transit region. When the charge neutrality in the region close to the heterojunction is satisfied further increasing in the applied potential causes a bending effect in the band structure. The transit region acts as a capacitor and the field increases inside it.

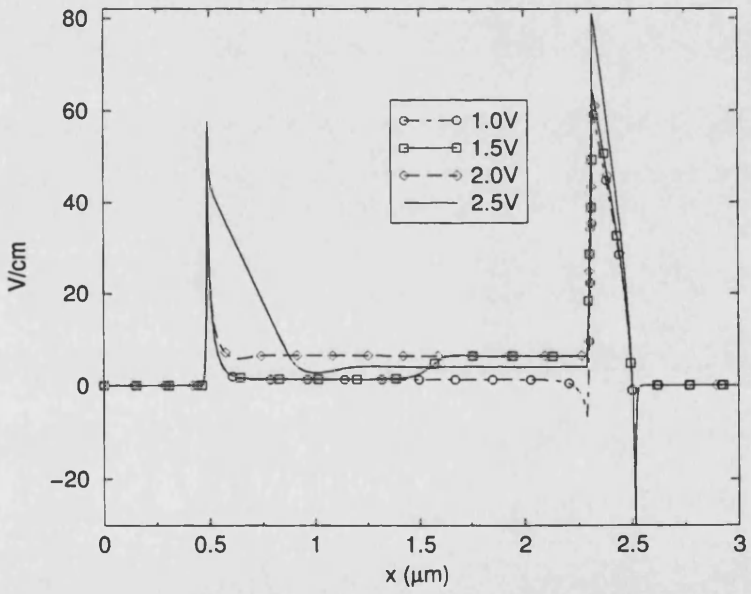

Figure 4.14: Electric field for different bias conditions as function of the position x in the device.

The results in figure 4.13 and figure 4.14 show that either a high bias or a heavily doped spike produces an accumulation of electrons near the anode contact. This can been explained observing that in both situations there is a non-equilibrium condition as far as the charge is concerned. A large number of electrons are injected in the transit region and accumulate near the anode. An accumulation-layer mode propagates from the cathode to the anode where it becomes stable as seen in section 4.2.1.

A series of Monte Carlo simulations has been carried out using the code developed by P. Briggs [20] to verify our drift-diffusion model results. Because of intrinsic difficulties in the Monte Carlo model, instead of the graded-gap injector we have substituted three layers of $Al_xGa_{1-x}As$ respectively with x=0.3, x=0.2 and x=0.1. The other parameters have been kept at the same value. Figure 4.15 shows the electric field in panel (a) and the electron density in panel (b) for a bias of 2 V. The solid line represents a device with a 20 nm spike inserted between injector and transit region. In accordance with the drift-diffusion results, the electric field has been tailored in the transit region. It is almost flat and above threshold, essential features for domain formation. On the other hand the electron density does not change significantly.

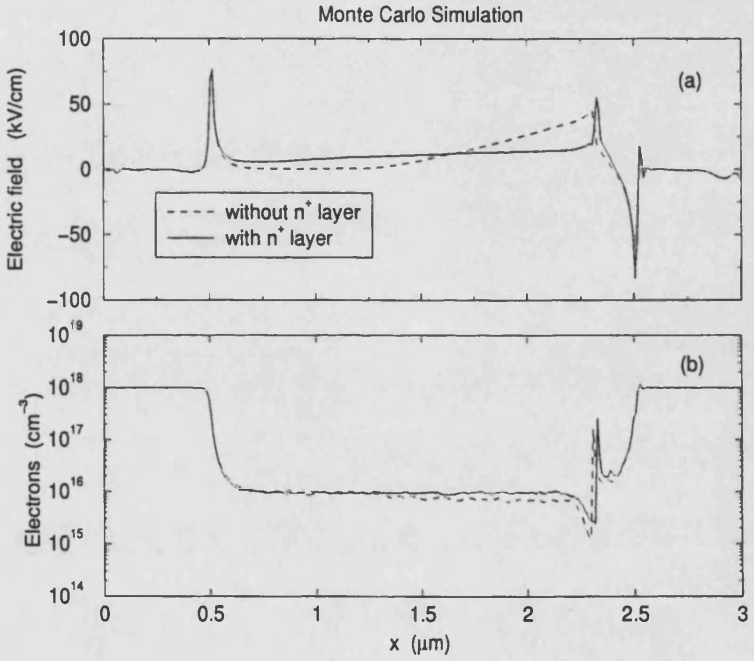

Figure 4.15: Monte Carlo simulation of the graded-gap Gunn diode with an without n^+ spike: (a) electric field and (b) density of electrons

Figure 4.16 shows the electron population percentage in the three valley bands for a structure with and without spike. A high percentage of electrons in the L-valley is found in high field regions of the device where electrons have enough energy to transfer in higher energy valleys. The difference between panels (a) and (b) is seen in the distribution of electrons in the transit region. A lower percentage of electrons in the Γ-valley, thus a higher percentage in the L-valley, is the proof of better injection by the launcher because of the n^+ layer.

The time-dependent simulation has shown that for a device without a spike the

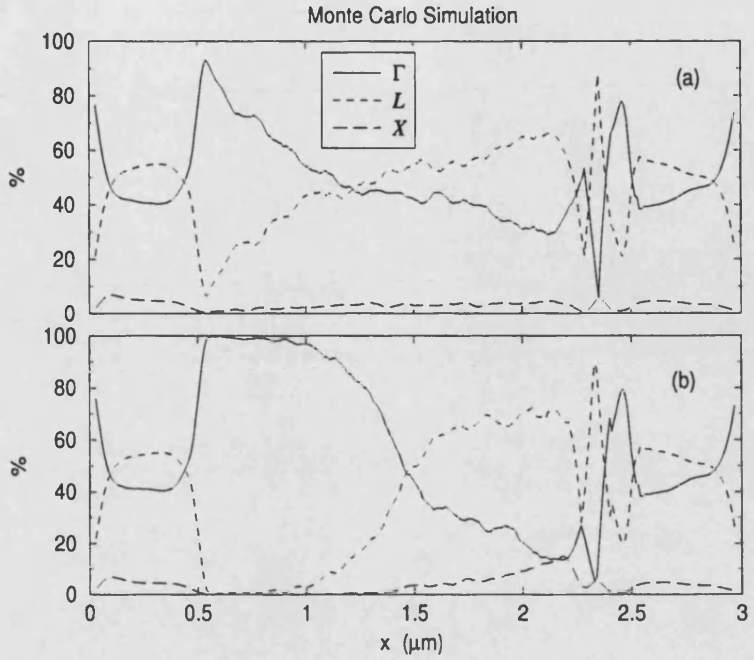

Figure 4.16: Monte Carlo simulation of the graded-gap Gunn diode: band occupation (a) without spike, (b) with spike

domain formation is largely reduced by the electric field profile and by the shortened transit region. Clear domains have been seen in a device with a spike and a launcher. In figures 4.17 and 4.18 clear self-sustained domain transport is present across the transit region. Experimental results on the same device structure give a frequency of ~50 GHz, i.e. a period of 20 ps for a 1.8 μm transit region length [22]. This has been verified by our simulations.

4.4 Conclusions

The preceding examples demonstrate the powerful engineering utility of simple hydrodynamic transport equations which govern a physical model of TEDs. This model is based on several phenomenological parameters derived from Monte Carlo particle simulations, and, therefore, the fundamental aspects of the semiconductor transport physics are built into simpler, tractable questions. Since the computer program is quite general, it should be noted that transport parameters from semiconductor materials other than GaAs, such as Si and InP, can be used as well. Several orders of magnitude reduction in computer costs and resources over the Monte Carlo method can be realised permitting efficient and systematic study

of device characteristics. Furthermore qualitative insights into important internal physical processes of carrier transport phenomena can be gained.

The program has been tested with Monte Carlo results and we have obtained a good agreement with the literature, in particular an almost perfect matching of results with previous works utilising a similar drift-diffusion approach [11].

We have concentrated our analysis mainly in the study of two particular devices. A simple n-layer structure has shown the importance of the boundary conditions for the mode of propagation of a Gunn diode. This has been useful in understanding of the dynamic of the second device: graded-gap AlGaAs Gunn diode. It has been seen again that it is possible to improve the performance of a Gunn diode by just modifying the cathode region. Monte Carlo simulations and experiments had allowed Spooner and Couch [23][14] to show the effects of insertion of a n^+ spike in graded-gap AlGaAs Gunn diode: the spike creates a charge sheet necessary for tailoring the electric field in the transit region and separate the injector from the transit region. These effects have been verified by our simulations. In addition, with our drift-diffusion model, we have been able for the first time to understand how the spike works and the limits of its use. We have proved that compensation of charge within the device is a key issue.

Currently, we are working for improving the speed and efficiency of the numerical technique and the computer algorithm. We are also writing up this work for publication in Electronics Letters.

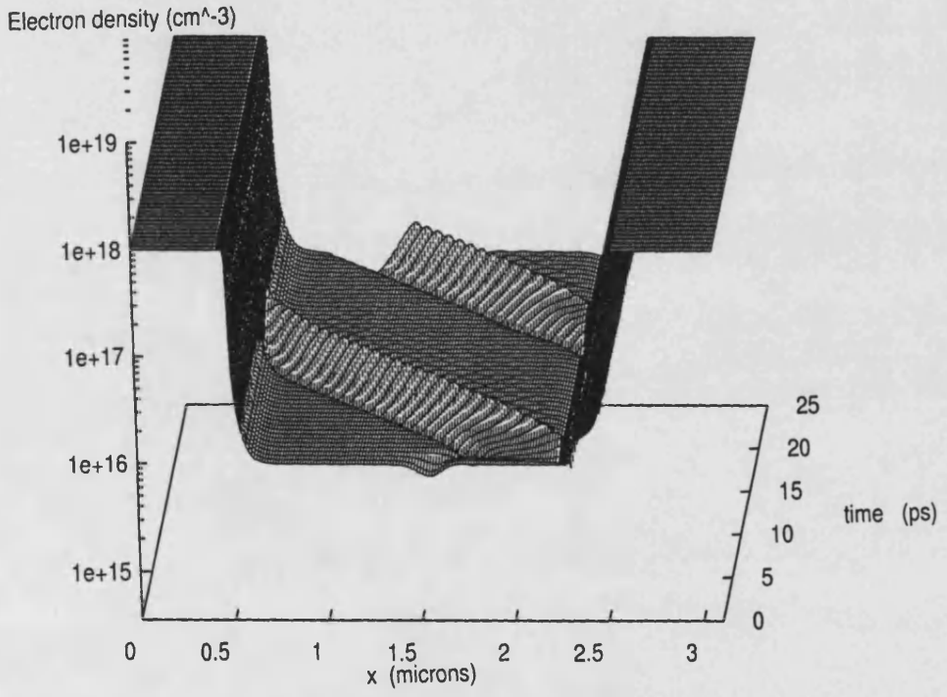

Figure 4.17: Electron density

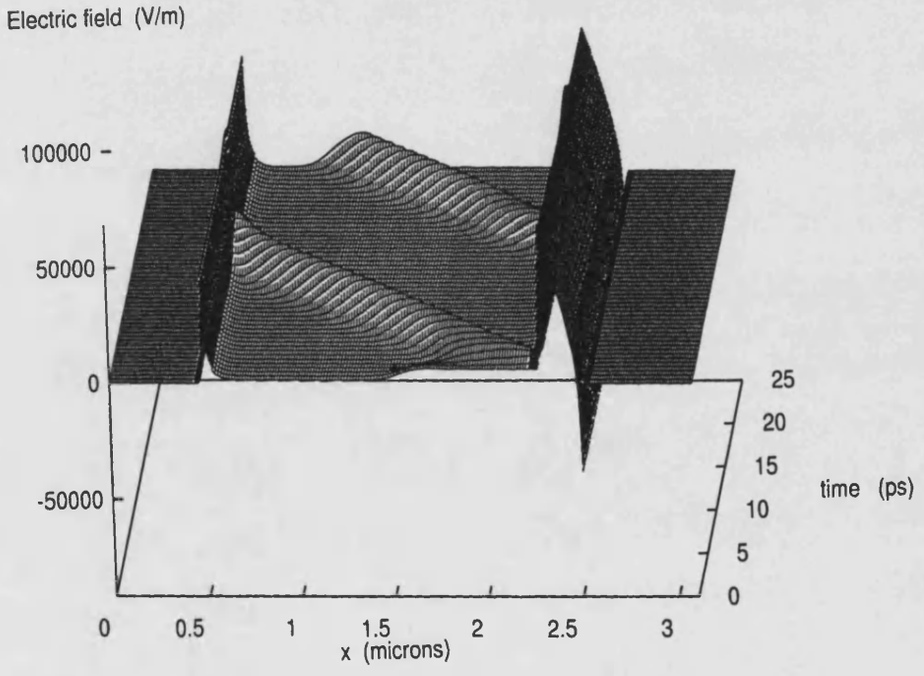

Figure 4.18: Electric field

References

[1] G. B. Tait and C. M. Krowne , *IEEE Transactions on Electron Devices*, **ED-3**, 223 (1988).

[2] R. B. Beall, S. J. Battersby, P. J. Grecian, S. Jones and G. Smith, *Electron. Lett.*, **25**, 871 (1989).

[3] R. Vaidyanathan and R. P. Joshi, *Electron. Lett.*, **27**, 1555 (1991).

[4] Z. Greenwald, D. W. Woodard, A. R. Calawa and L. F. Eastmann, *Solid-State Electron.*, **31**, 1211 (1988).

[5] N. R. Couch, H. Spooner, P. H. Beton, M. J. Kelly, M. E. Lee, P. K. Rees and T. M. Kerr, *IEEE Electron Device Lett.*, **EDL-10**, 288 (1989).

[6] M. R. Lakshminarayana and L. D. Partain, *IEEE Trans. Electron Devices*, **ED-27**, 546 (1980).

[7] R. B. Robrock, *IEEE Trans. Electron Devices*, **ED-16**, 647 (1969).

[8] H. Thim, *Proceeding IEEE*, **59**, 1285, (1971).

[9] P. Gueret and M. Reiser, *Appl. Phys. Lett.*, **20**, 60 (1972).

[10] S. M. Sze, *"Physics of Semiconductor Devices,"* John Wiley & Sons, 1981.

[11] G. B. Tait and C.M. Krowne, *Solid-State Electronics*, **30**, 1025 (1987).

[12] S. Hutchinson, J. Stephens, M.J. Kelly and M. Carr, *Electronics Letters*, **32**, 851 (1996).

[13] N. R. Couch, H. Spooner, M. J. Kelly and T. M. Kerr, *Solid-State Electronics*, **32**, 1685 (1989).

[14] N. R. Couch, A. P. Long, P. H. Beton and M. J. Kelly, *Electronics Letters*, **24**, 434 (1988).

[15] M. R. Friscourt, P. A. Rolland and M. Pernisek, *IEEE Electron Device Letters*, **6**, 497 (1985).

[16] M. S. Lundstrom and R. J. Schuelke, *IEEE Transactions on Electron Devices*, **30**, 1151 (1983).

[17] K. Horio and H. Yanai, *IEEE Transactions on Electron Devices*, **37**, 1093 (1990).

[18] S. Adachi, *Journal of Applied Physics*, **58**, R1 (1985).

[19] H. Thim, *Electronic Letters*, **7**, 246 (1971).

[20] P. J. Briggs, *Transport Effects in Heterojunction Bipolar Transistors*, Ph.D. thesis, University of East Anglia (1997).

[21] G. Khrenov, V. Ryzhii and O. Kartashova, *Solid-State Electronics*, **37**, 1789 (1994).

[22] M. A. Harry, M. J. Kelly, G. M. Dunn, M. J. Kearney, J. Stephens and M. Carr, *Proceeding IEEE*, **52**, 162 (1998).

[23] H. Spooner and N. R. Couch, *GEC Journal of Research*, **7**, 34 (1989).

Chapter 5

Nitrides

5.1 Introduction

There are many areas where conventional III-V semiconductors cannot be used. Short wavelength light emitters are required for full colour display, laser printers, high density information storage, and under water communication. High temperature and high power transistors are needed for automobile engines, future advanced power distribution systems, all electric vehicles and avionics. Si and conventional III-V semiconductors are not suitable for designing and fabricating optoelectronic devices in the violet and blue region of the spectrum. Their band gaps are not sufficiently large. GaAs based electronic devices can not be used at high temperatures. III-nitrides are particularly suitable for applications in these areas. The band gaps of the III-nitrides are large and direct. The band gaps values are 1.9 eV for InN, 3.4 eV for GaN, 6.2 eV for AlN [1]. Because of their wide band gaps and strong bonds, they can be used for violet, blue and green light emitting devices and for high temperature transistors.

InN and AlN can be alloyed with GaN. This allows tuning of the band gaps and emission wavelengths. Using AlGaN or GaN as the barrier and cladding layers, and GaN or InGaN as active layers, quantum wells and superlattices can be fabricated. This is important because the heterostructure technology used in the fabrication of superlattices is essential for fabricating modern devices. InGaN quantum wells are indispensable for light emitting devices because incorpora-

tion of small In in the active GaN layer increases the luminescence efficiency considerably. ZnSe based semiconductors also have band gaps suitable for short-wavelength optoelectronic devices but the bond strength in the II-IV wide band gap semiconductors is low. Doping of the III-nitrides with both donors and acceptors has been accomplished. The Monte Carlo simulated peak velocity in GaN is close to $3{\cdot}10^7$ cm/s and the saturation velocity is $1.5{\cdot}10^7$ cm/s [2]. Both values are considerably higher than the values for Si and GaAs. The low field mobilities of electrons and holes in GaN are reasonable although they are not as high as in GaAs. The highest reported room temperature electron mobility in GaN epilayer is 900 cm^2/Vs [3]. III-nitrides devices can work not only at much higher temperatures but also in hostile environments thanks to several properties one of which is the good thermal conductivity [3]. GaN has high breakdown fields, $3{\cdot}10^6$ V/cm [4].

Despite progress in device fabrication, there are still substantial problems concerning growth optimisation and insight into the fundamental mechanisms is rather shallow. Even the atomic and electronic structures of both the cubic and hexagonal crystals are not well known. The very specific properties of the III-nitrides lead to a number of unique physical properties, which are not just important in device fabrication but also interesting from a theoretical point of view. For example, the relatively small size of nitrogen, compared to Ga or In, in these compounds leads to a vast range of scenarios when heterointerfaces are considered.

In this chapter first we outline the main characteristics of two of the three principal nitride materials, GaN and InN, with particular attention to their alloy InGaN. AlN and AlGaN will not be considered in this work but much information on these materials can be obtained from [1][5]. We then describe the problem of the strain inside nitride heterostructure and we will present many issues related to spontaneous and strain-induced polarization. Finally we will describe the functionality of a GaN/InGaN LED and its limitations.

5.2 Gallium Nitride

More than a quarter century has elapsed since the first demonstration of stimulated emission in GaN under optical pumping [6]. Progresses have been slow because of intrinsic difficulties linked to the growth of GaN and the lack of appropriate substrates. For almost two decades, *p*-doping of GaN was not feasible. With the achievement of *p*-doping by an annealing process [7][8] devices based on GaN were demonstrated. Since the technological breakthroughs by Nichia like a blue laser diode working at room temperature[9], in cw mode and lasting 10,000 hours, gallium nitride has been the subject of intensive research activity worldwide. Devices based on GaN would have not been possible without the emergence of new approaches to the heteroepitaxy of GaN leading to layers exhibiting lower dislocation densities than those grown using conventional heteropitaxy.

5.2.1 Growth

The high melting point of GaN rules out the application of the standard crystallisation techniques like Czochralski or Bridgman growth from the stoichiometric melt as for Si. GaN crystals are usually grown in the form of thin epitaxial layers on a substrate, by low temperature methods like MBE or MOVPE. During the growth process high N_2 pressure is required for GaN stability at high temperatures. A temperature in excess of 2500 K and a N_2 pressure on the order of tens of kbars are necessary for the melting of GaN. The solubility of nitrogen in liquid Ga on the order of 1% can be reached by the application of N_2 pressure of 10-20 kbar at temperatures 1500-1600 °C .This allows the growth of high quality GaN single crystals from the solution in temperature gradients [10].

Sapphire (Al_2O_3) is the most extensively used substrate for growth of the III-nitrides. Large area, good quality crystals of sapphire are easily available at low cost. The crystal orientations of sapphire and GaN (grown on *c*-plane [0001] sapphire) are parallel, but the unit cell of GaN is rotated by 30° about the *c* axis with respect to the sapphire unit cell [1]. The poor quality is the result of a large mismatch in the lattice constants (~15%) and in the coefficients of thermal expansion (~80%) of GaN and sapphire. Epylayers grown directly on sapphire are strained and with a high density of defects. Good quality films can be grown

instead by first growing a low temperature GaN buffer layer [11]. The buffer layer provides the high density of nucleation centres and promotes the lateral growth of the main epilayer. The properties of the GaN epilayer are strongly influenced by the thickness and growth temperature of the buffer layer.

Among other substrates we mention SiC that has several advantages over sapphire. The main properties are that its lattice mismatch with GaN is only 3.5%, and it has a good electrical conductivity. A big drawback is the cost of producing SiC. To accomplish integration of III-nitride devices with Si electronics, considerable work has been done also on the growth of GaN on the Si substrate [12]. Unfortunately, because of the large mismatch, large number of defects, such as dislocations, twins and stacking faults are predominant.

A significant effort has been put on the reduction of the threading defect density in GaN since these defects are known to jeopardise the device functionality. Using lateral epitaxy overgrowth (ELOG), large areas of low defect density GaN are produced [9].

5.2.2 Crystal structure

GaN usually crystallises in the hexagonal wurtzite (WZ) structure, with two formula units per unit cell (4 atoms per cell) and a molecular weight of 83.728 g/mol. The space group symmetry is C_{6v}^4 ($P6_3mc$) and the point group symmetry is C_{6v} (6mm). Strain and defects may distort the lattice constants from their intrinsic values, and thus there is a wide dispersion in reported values. For WZ-GaN at room temperature, lattice parameters of a_0=3.1892±0.0009 $\dot{A}$ and c_0=5.1850±0.0005 $\dot{A}$ [13][14] are generally accepted. For the zinc blende (ZB) and rocksalt structures, refer to [5].

5.2.3 Mechanical properties

In 1997 investigations of the elastic constants of WZ-GaN were carried out by Savastenko and Sheleg on powdered GaN crystals using X-ray methods [15]. The measured elastic constants are shown in the first row of table5.2.3. Table 5.2.3

reports the values of elastic constants as found in the literature. The Poisson's ratio (last column in table 5.2.3) is defined as $\mu_{<0001>} = (\Delta a/a_0)/(\Delta c/c_0)$ and relates the deformation of the unit cell along the in-plane a axis and the deformation along the $< 0001 >$ c axis. It can be estimated from the elastic constants using the following equation:

$$\frac{\nu}{1-\nu} = -\frac{C_{13}}{C_{33}}. \tag{5.1}$$

Section 5.4.4 will consider with more details the Poisson's ratio for a nitride heterostructure.

Reference	C_{11}	C_{12}	C_{13}	C_{33}	C_{44}	$\nu_{<0001>}$
Savastenko('78)[15]	29.6	13.0	15.8	26.7	2.41	-1.45
Polian('96)[16]	39.9	14.5	10.6	39.8	10.5	-0.363
Takagi('96)[17]	37.4	10.6	7.0	37.9	10.1	-0.227
Wright('97)[18]			10.3	40.5		-0.341
Yamaguchi('97)[19]	36.5	13.5	11.4	38.1	10.9	-0.427
Schwarz('97)[20]	37.7	16.0	11.4	20.9	8.14	-1.2
Deger('98)[21]	37.0	14.5	11.0	39.0	9.0	-0.393
Deguchi('99)[22]	37.3	14.1	8.0	38.7	9.4	-0.261

Table 5.1: Room temperature elastic constants (C_{ij} in dyn/cm^2) and Poisson's ratio (ν) of WZ-GaN

Analysing the composition dependence of emission energy of pseudomorphically strained InGaN layers with In content up to 0.2, Stepanov et al.[23] found that all the experimental data yeld essentially the same results when a value of the Poisson's ratio of ν=0.205 is chosen. This result, that does not depend on deriving the Poisson's ratio from the elastic constants, is in good agreement with the most recent calculation of Deguchi et al.[22].

5.2.4 Thermal Properties

Thermal expansion of single crystal WZ-GaN has been studied in the temperature range from 300 K to 900 K, or from 80 K to 820 K. Maruska and Tietjen [13] reported that the lattice constant a changes linearly with the temperature with a coefficient of thermal expansion of $5.59 \cdot 10^{-6}$ K^{-1}. Meanwhile, the expansion of lattice constant c shows a superlinear dependence on temperature. Other thermal and thermodynamic properties for WZ-GaN have been tabulated by Elwell et al.[24].

5.2.5 Band Structure

The band structure of WZ GaN has been investigated by several groups [25][26]. The calculated values of the band energy levels [26] are shown in figure 5.1.

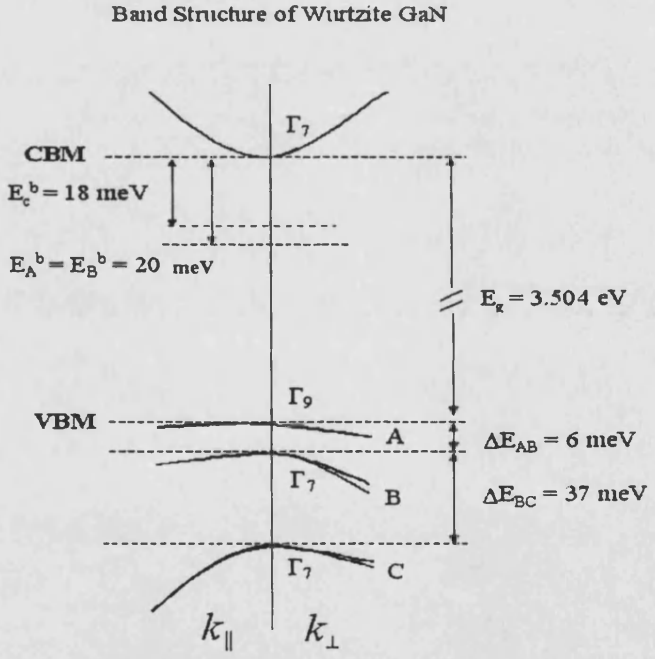

Figure 5.1: Calculate values of the band energy levels of wurtzite GaN [26]

The degeneracy of the valence band is lifted due to crystal field and spin-orbit coupling. The valence band splits into three bands designated as A, B and C bands (see figure 5.1). The symmetry of the A band is Γ_9 and that of the B and C bands is Γ_7. In the A, B and C bands, the average hole mass is given by

m_h/m_0=0.6 while in conduction band the electron effective mass is m_e/m_0=0.18 [26]. The hole mass has been obtained as average value of light and heavy hole effective masses. We have not considered any anisotropy of the bands or band splitting due to the strain in the lattice. An approach that will includes these effects requires a complete band structure calculation and it will be necessary for quantitative predictions.

5.2.6 Optical Properties

At low temperatures, a typical near band-edge PL spectra of GaN is dominated by the BX peak($\sim$3.49 eV) due to excitons bound to impurities. Since the GaN is generally n-type the BX peak is most likely due to the excitons bound to neutral donors. Peaks due to free excitons involving holes in the A and the B valence bands (see figure 5.1) become comparable with BX peak and dominate for T >50 K. The strength of the PL is sensitive to the quality of the crystal. Nonradiative recombination is promoted by multiphonon processes, Auger recombination, recombination at defects and recombination at surfaces. Chen et al.[25] have calculated the binding energies ΔE_A^b, ΔE_B^b and ΔE_C^b of the A, B and C free excitons bands. The calculated values are shown in figure 5.1.

PL peaks are also observed on the low energy side of the near band-edge spectra. The main features are a yellow band at about 2.5 eV and donor-acceptor pair (DAP) bands. The origin of the yellow band is not completely known. The vibronic structure of the yellow band shows the strong electron-phonon coupling. Therefore, this band may be associated with deep levels derived by point defects such as Ga or N vacancies.

5.2.7 Electrical Properties

As grown GaN is always n-type with carrier concentration $\sim 10^{16}$ cm^{-3}. Two models exist to explain the n-type conductivity in GaN epilayers grown without intentional doping. These models also explain the yellow emission band seen in all the experimental results. According to one model, the nitrogen vacancies act as shallow donors and give rise to n-type conductivity in nominally undoped

GaN epilayers [27]. According to the other model, residual impurities (Si, Ge or O) give rise to the n-type conductivity. These impurities and Ga vacancies are also the centres of the yellow emission [28]. Until now, neither of the two model has been excluded. Other experiments suggest that the centres that give rise to yellow band and n-type conduction are different in samples grown by different techniques and under different experimental conditions [1].

Both Ge and Si substitute Ga site in GaN and form shallow donors with an activation energy in the range 12-22 meV. The room temperature mobility measured by Nakamura et al.[29] is 600 cm^2/Vs for an electron concentration of $\sim 3 \cdot 10^{16}$ cm^{-3}, and decreases slowly with the increase in carrier concentration, reaching a value of about 100 cm^2/Vs at a carrier concentration of $3 \cdot 10^{18}$ cm^{-3}. For an electron concentration of $1 \cdot 10^{18}$ cm^{-3}, the mobility is 250 cm^2/Vs.

Name	Symbol	Value	Unit
Energy gap	E_g	3.39	eV
Relative dielectric constant	ϵ_r	8.9	
Electron mobility	μ_n	600	cm^2/Vs
Hole mobility	μ_p	15	cm^2/Vs
Fraction of electron effective mass	m_e	0.18	
Fraction of hole effective mass	m_h	0.6	
Electron recombination time	τ_n	$1 \cdot 10^{-7}$	s
Hole recombination time	τ_n	$7 \cdot 10^{-9}$	s
Ionisation donor energy	E_D	$20 \cdot 10^{-3}$	eV
Ionisation acceptor energy	E_A	$150 \cdot 10^{-3}$	eV
In-plane lattice constant	a	3.189	Å
c-axis lattice constant	c	5.187	Å
Piezoelectric tensor component	e_{33}	0.73	C/m^2
Piezoelectric tensor component	e_{31}	-0.49	C/m^2
Poisson's ratio	ν	0.205	
Equilibrium Polarization	P_{eq}	-0.029	C/m^2

Table 5.2: GaN material parameters [1]

Doping GaN with acceptors to obtain a high concentration of holes has been a difficult problem. Attempts to dope GaN with Mg acceptors were not successful because Mg was always compensated by the background n-type defects or was passivated with hydrogen atoms. It was discovered by Amano et al.[7] that Mg-doped GaN can be converted into p-type GaN by low energy electron beam irradiation (LEEBI). It was shown that Mg is an acceptor with an energy of activation equal to 160 meV. Nakamura et al.[30] have measured the Hall mobility

of holes as a function of carrier concentration in LEEBI treated Mg-doped GaN epilayers grown on GaN buffers. For an hole concentration of $5{\cdot}10^{17}$ cm^{-3}, the mobility is 15 cm^2/Vs and it decreases to 5 cm^2/Vs when holes have a density of $1{\cdot}10^{19}$ cm^{-3}.

5.3 Indium Nitride

In the absence of good quality single crystal samples, the physical properties of indium nitride have been measured on non ideal thin films, typically ordered polycrystalline with crystallites in the 50 nm to 500 nm range. Structural, mechanical and thermal properties have not been yet reported for epitaxial films on lattice-matched substrates. The high density of nitrogen vacancies affects many properties including lattice parameters. Material not grown under ultra-high vacuum conditions is susceptible to high levels of oxygen incorporation which may affect physical properties.

5.3.1 Crystal structure

Indium nitride normally crystallises in the wurtzite (hexagonal) structure. Table 5.3.1 lists values of basal a and perpendicular c lattice constants. The ratio c/a approaches the ideal value of 1.633 only in the case of growth with precautions taken to reduce N vacancies [31].

Reference	a	c	c/a
Tansley('86)[31]	0.35480	0.576	1.6234
Kubota('89)[32]	0.3540	0.5705	1.612
Strite('93)[33]	0.36	0.574	1.59

Table 5.3: Lattice parameters of WZ-InN

It is also noteworthy that uncertainties in c are about ten times greater then uncertainties in a. This may be a function of nitrogen deficiency, since N atoms are close packed in (0001) planes and high vacancy densities may preferentially

shrink the lattice in the perpendicular direction, parallel to c. Estimates of the principal transverse and longitudinal elastic constants c_t and c_l are respectively $4.42 \cdot 10^{11}$ and $2.65 \cdot 10^{12}$ dyn cm^{-2} [5].

5.3.2 Mechanical properties

The values of the five distinguishable second-order elastic moduli in a hexagonal crystal are C_{11}, C_{12}, C_{13}, C_{33} and C_{44} are shown in table 5.3.2. There are few reports of measured and calculated values in literature. However, since each elastic constant depends principally on the lattice constants [34] which vary by only about 10% across the nitrides, values for GaN may be used as a first approximation[5].

Reference	C_{11}	C_{12}	C_{13}	C_{33}	C_{44}	$\mu_{<0001>}$
Sheleg('79)[35]	19.0	10.4	12.1	18.2	9.9	1.983
Reeber('01)[36]	22.3	11.5	9.2	22.4	4.8	0.697

Table 5.4: Room temperature elastic constants (C_{ij} in dym/cm^2) and Poisson's ratio ($\mu_{<0001>}$) of WZ-InN

5.3.3 Electrical Properties

Hall measurements show that InN films are always n-type with a high concentration of electrons [5]. The electron concentration decreases when the growth temperature of InN films is increased. This observation argues against the nitrogen vacancies being responsible for the background n-type conductivity. At high growth temperatures, the nitrogen atoms should evaporate and the concentration of nitrogen vacancies increase, but the measured mobility increases with the growth temperature. Values around 250 cm^2/Vs were reported by Hovel et al.[37] at room temperature. Lower values of 35-50 cm^2/Vs were measured by Trainor et al.[38]. Measurements of the piezoelectric constant has not been reported, but its dependence on the dielectric constants ε_r and e_{14} [39] suggests that values of about 50% of those found in GaN can be used [5].

Name	Symbol	Value	Unit
Energy gap	E_g	1.89	eV
Relative dielectric constant	ϵ_r	15.3	
Electron mobility	μ_n	3200	cm^2/Vs
Hole mobility	μ_p	30	cm^2/Vs
Fraction of electron effective mass	m_e	0.11	
Fraction of hole effective mass	m_h	0.4	
Electron recombination time	τ_n	$7{\cdot}10^{-8}$	s
Hole recombination time	τ_n	$7{\cdot}10^{-10}$	s
Ionisation donor energy	E_D	$5{\cdot}10^{-2}$	eV
Ionisation acceptor energy	E_A	$5{\cdot}10^{-2}$	eV
In-plane lattice constant	a	3.533	Å
c-axis lattice constant	c	5.705	Å
Piezoelectric tensor component	e_{33}	0.97	C/m^2
Piezoelectric tensor component	e_{31}	-0.57	C/m^2
Poisson's ratio	ν	0.205	
Equilibrium Polarization	P_{eq}	-0.032	C/m^2

Table 5.5: Summary of InN material parameters as reported in [5]

5.4 InGaN/GaN heterostructures

InGaN-based LEDs and LDs have great commercial potential due to their ability of working in the short wavelength region (green emission). The active region of these devices, in which the light is being generated, is an InGaN/GaN or InGaN/AlGaN quantum well. Even though devices based on this quantum well structure are already being mass-produced and are available in the market, the mechanisms by which this light is generated are poorly understood. The reason of this lack of understanding can be found in the large number of complex phenomena that occur when InGaN layer is grown on GaN or AlGaN substrates. In this section we will consider few of these phenomena that have direct consequences on our work.

5.4.1 Misfit Strain and Critical Thickness

A pseudomorphic epitaxial layer of $In_xGa_{1-x}N$ can be grown on a GaN substrate. If the thickness h of the layer is smaller than a certain thickness, known as the critical thickness h_c, the misfit between the layer and the substrate is ac-

commodated by a tetragonal compression of the $In_xGa_{1-x}N$ layer. The strain in the layer is homogeneous and is known as *misfit strain.* The in-plane lattice constant a of the InGaN layer becomes the same as that of the GaN substrate. Such homogeneously strained layers are known as *pseudomorphic* strained layers. When $h > h_c$, misfit strain begins to relax by the introduction of misfit dislocations. Akasaki and Amano [40] investigated the critical thickness of $In_xGa_{1-x}N$ layer grown on relaxed GaN layers. The concentration x of In was in the range $0.05< x <0.2$. The value of critical thickness was ~400 nm.

Strain has a profound effect on many properties of the epilayers. For example, the band structure and electronic properties of III-nitrides are considerably modified by strain [1]. A typical example is the problem of the band offset. Knowledge of the band offset at the heterojunction is vital for determining the degree of confinement of the carriers in quantum well and quantum dot structures. The value of the band offset for a given $A-B$ heterojunction (where B is the substrate and A is grown on the top of B) is not unique. It depends on whether the pseudomorphic epilayer A is grown on the relaxed layer B or of the strained epilayer B grown on the relaxed layer A. The difference arises because strain gives rise to piezoelectric fields (see section 5.4.2) [41]. Martin et al.[42] report that for pseudomorphic InN grown on the top of relaxed GaN, the ratio between the band offset in conduction band and the difference in energy gap between the two material is $\Delta E_c/\Delta E_g$=0.57.

5.4.2 Piezoelectricity

III-nitrides lack inversion symmetry and exhibit piezoelectric effects when strained along [0001]. The stress-strain relations for a hexagonal crystal are given by a 6×6 matrix. However if the crystal is strained in the (0001) plane and is free in the [0001] direction, there are only three nonvanishing strain components: $\varepsilon_{11} = \Delta a/a_0$, $\varepsilon_{22} = \Delta b/b_0$ and $\varepsilon_{33} = \Delta c/c_0$. If the strain is biaxial, a=b and so the strains $\varepsilon_{11} = \varepsilon_{22}$. The piezoelectric coefficients are almost an order of magnitude larger than in traditional III-V semiconductors as shown in tables 5.2.7 and 5.3.3.

Moreover, wurtzite is the structure with the highest symmetry compatible with the existence of spontaneous polarization. Spontaneous polarization in III-

nitrides was only recently fully understood and it has been studied by many groups [43][44][45][46][47]. Tables 5.2.7 and 5.3.3 report the established values of the spontaneous polarization for GaN and InN.

In chapter 2 section 2.9 we have explain a mathematical method to calculate strain-induced polarization. In chapter 6 we will describe in further detail the difficulty of deducing the piezoelectric field in InGaN quantum wells.

5.4.3 Energy gap and bowing parameter of $In_xGa_{1-x}N$

The band gap of $In_xGa_{1-x}N$ alloys depends on the alloy composition and also on temperature. At low temperatures (~10 K) the measured band gap is given by the following expression [48]:

$$E_g(x) = 3.5 - 2.63x + 1.02x^2. \tag{5.2}$$

The band gap decreases with temperature and at 295 K it is smaller by about 0.006 eV for all values of x. The bowing parameter b is defined as

$$E_g(x) = E_{g,InN}x + E_{g,GaN}(1-x) - bx(1-x), \tag{5.3}$$

and its value derived from equation 5.2 is b=1.02 eV. However, other authors give different values: b=3.8 eV [49], b=1.89 eV [50], b=2.7 eV [23], b= 3.2 eV [40]. The PL and absorption techniques give different values of the band gap and of the bowing parameter. The measured values of the bowing parameter also depend on the technique for measuring the band gap. In this case uncertainties arise because the In concentration is not known accurately and is nonuniform. In section 5.4.4 we will present a procedure [51] for calculating the indium composition from the measurements of lattice constants. From that calculation it is clear that the knowledge of the elastic constants and therefore of the Poisson's ratio is critical in order to calculate the value of the bowing parameter.

5.4.4 Chemical composition and lattice parameters of semiconductor alloy

The determination of the chemical composition by x-ray diffraction (XRD) is based on a unique relationship between the lattice parameter and the chemical composition, which is generally assumed to be linear for semiconductor alloy systems (Vegard's rule). Under this assumption, in a two-component alloy system A_xB_{1-x}, the A content x and B content $(1-x)$ are related to the relaxed lattice parameters a_0 and c_0 of the hexagonal unit cell by following expressions [51]:

$$c_0(L) = xc_0(A) + (1-x)c_0(B) \tag{5.4}$$

$$a_0(L) = xa_0(A) + (1-x)a_0(B) \tag{5.5}$$

where L indicates the A_xB_{1-x} alloy film and A and B the components. For example, in the alloy $In_xGa_{1-x}N$, x is the InN content and $1-x$ the GaN content. The chemical composition, i.e. the component A mole fraction x is

$$x = \frac{c_0(L) - c_0(B)}{c_0(A) - c_0(B)}. \tag{5.6}$$

The functional group-III nitride alloy films are usually not grown directly on the sapphire substrates, but on a GaN buffer layer to improve the crystalline quality. Between the group III-nitride alloy film and the GaN buffer, a nearly perfect in-plane orientation is established. The buffer forms a virtual substrate of relaxed GaN. Below a critical thickness, thin films exhibit pseudomorphic growth, where the basal planes of the hexagonal unit cells coincide. Therefore, in the basal plane of the heterostructure, a biaxial strain is produced which causes a distortion of the hexagonal unit cell and a change of the c/a ratio. The distortion of the unit cell can be described by the generalised Hooke's law using the tensor of the elastic moduli of hexagonal crystal systems. In the epitaxial systems under consideration, the c axis of the unit cell is parallel to the z or the direction labelled with the number 3 (normal to the substrate surface), which allows a simplification of the tensor of the elastic moduli for hexagonal unit cells.

In the biaxial strain case, the stress normal to the film vanishes σ_{33}=0 and the following ratio between the strain ε_{33} in the growth direction and the strain ε_{11} in the basal plane is obtained

$$\varepsilon_{33} = -\frac{C_{13}}{C_{33}}(\varepsilon_{11} + \varepsilon_{22}) \tag{5.7}$$

where σ_{ij} is the stress tensor, C_{ij} the tensor of the elastic moduli in Voigt's notation. The components of this tensor have been defined already as elastic constants and their value for GaN and InN are reported in table 5.2.3 and 5.3.2, respectively. ε_{kl} is the strain tensor, which is isotropic ε_{11}=ε_{22} in the basal plane.

In pseudomorphic films as well as in partially relaxed films, the hexagonal crystal structure is maintained. With the definitions for the strains $\varepsilon_{33} = (c(L) - c_0(L))/c_0(L)$ and $\varepsilon_{11} = (a(L) - a_0(L))/a_0(L)$, equation 5.7 can be written as

$$\frac{c(L) - c_0(L)}{c_0(L)} = -2\frac{C_{13}(L)}{C_{33}(L)}\frac{a(L) - a_0(L)}{a_0(L)} \tag{5.8}$$

where $c(L)$ and $a(L)$ are the lattice parameters of the strained InGaN film and $c_0(L)$ and $a_0(L)$ are the corresponding relaxed lattice parameters. The strained lattice parameters on the film $c(L)$ and $a(L)$ are measured directly by XRD. For pseudomorphic growth, $a(L) = a_0(S)$, where S indicates the substrate or the GaN buffer, as in our case. The elastic moduli C_{13} and C_{33} depend on the chemical composition of the alloy.

For Poisson's ratio as a first-order approximation, a linear relationship is applied

$$\nu(L) = x\nu(A) + (1 + x)\nu(B). \tag{5.9}$$

Substitution of $a_0(L)$, $c_0(L)$ and $\nu(L)$ in 5.8 results in the following cubic equation for x

$$Px^3 + Qx^2 + Rx + S = 0 \tag{5.10}$$

where

$$P = (\nu(A) - \nu(B))(a_0(A) - a_0(B))(c_0(A) - c_0(B)) \tag{5.11}$$

$$\begin{aligned} Q &= (1 + \nu(B))(a_0(A) - a_0(B))(c_0(A) - c_0(B)) + \\ &\quad (\nu(A) - \nu(B))[(a_0(A) - a_0(B))c_0(B) + \\ &\quad (a_0(B) - a(L))(c_0(A) - c_0(B))] \end{aligned} \tag{5.12}$$

$$\begin{aligned} R &= (a_0(A) - a_0(B))[(1 + \nu(B))c_0(B) - c(L)] + \\ &\quad (c_0(A) - c_0(B))[(1 + \nu(B))a_0(B) - \nu(B)a(L)] + \\ &\quad (\nu(A) - \nu(B))(a_0(B) - a(L))c_0(B) \end{aligned} \tag{5.13}$$

$$S = (1 + \nu(B))a_0(B)c_0(B)) - \nu(B)a(L)c_0(B) - a_0(B)c(L). \tag{5.14}$$

Equation 5.10 describes the general case, in which the elastic moduli vary with the chemical composition, i.e. $C_{13}(A) \neq C_{13}(B)$ and $\nu(A) \neq \nu(B)$. The solution of this equation gives the chemical composition, i.e. the In mole fraction x in the case of $In_xGa_{1-x}N$. There are three solutions, but only one physically meaningful value of x in the interval $0 \leq x \leq 1$. With this mole fraction, using 5.4, 5.5 and 5.9, the strains ε_{33} and ε_{31}, as well as Poisson's ratio $\nu(L)$ can be calculated.

5.4.5 Structural defects

The GaN-based III nitride heterostructures are found to contain characteristic one-dimensional (edge, mixed and screw dislocation) and two-dimensional (stacking faults and domain boundaries) defects. These defects act as nonradiative recombination centres. Lester et al.[52] measured the dislocation concentration in the active layers in high efficiency commercial LEDs of GaN devices as $\sim 10^{10}$ cm^{-2}. Until recently the density of the threading dislocations present in practi-

cally all the high performance LEDs and laser diodes was 10^8-10^{12} cm^{-2}. The main reason for this high density is the large lattice mismatch between the III-nitrides and the sapphire substrate. In conventional III-V semiconductor devices, a dislocation density of $>10^6$ cm^{-2} would be fatal for light emission as the dislocations are nonradiative recombination centres and reduce the optical output. The heat generated by recombination at the dislocations induces motion and multiplication of defects and very soon the devices burn out or stop emitting light. In spite of this large dislocation density, the efficency of the GaN-based LEDs is higher than AlGaAs and AlInGaP-based LEDs and light emitters. Recent work shows that dislocation do reduces considerably the light output from the GaN-based LEDs and they increase the leakage current [53]. Jain et al.[54] believes that dislocations are trapping recombination centres that can cope with a limited number of recombinations per unit time. As the carrier density increases, the loss of carriers by recombination increases in the beginning but saturates at a constant value for large injected carrier density. For sufficiently large density of the injected carriers, the fraction lost by the recombination at dc becomes negligible.

5.4.6 Optical Properties of InGaN

Experiments show that In concentration in InGaN alloys is not uniformly distributed [55]. It is suggested that the emission in the InGaN material is from localised excitons. The origin of these localised states is the In-rich regions due to fluctuations in the alloy composition. In-rich regions act as quantum dots and give rise to efficient radiative recombination. Other authors suggest that the main radiative recombination mechanism is dominated by strong polarization fields induced by the strain in the layer [56]. However, there is much disagreement between the researchers involved in this problem as to which recombination mechanism is responsible for light emission in these devices. In chapter 6 we will have to face this issue and we will try to suggest a solution using the model we have developed in chapter 2.

5.5 Conclusion

We have presented a brief summary of the physics of III-nitrides with particular attention to GaN and InN. We have shown that the knowledge of many parameters for nitrides is far from satisfactory, or at least less than the other III-V semiconductors. Even after more than 30 years of research, III-nitrides remain *immature* semiconductors. A major consequence of this is the impossibility of separating optical or electrical transport issues from the growth problems. Device simulation has to be able to include new physical phenomena and has to be flexible enough to include effects that are material or even sample dependent. Almost all the parameters reported have been used in our model or have been kept in consideration for a future development of the code.

References

[1] S. C. Jain, *Appl. Phys. Rev.*, **87**, 965, (2000).

[2] B. Gelmot, K. Kim and M. Shur, *J. Appl. Phys.*, **74**, 1818 (1993).

[3] S. C. Binari and H. C. Dietrich, *GaN and Related Materials*, edited by S. J. Pearton, Gordon and Breach, New York, 509 (1997).

[4] J. I. Pankove, *GaN and Related Materials*, edited by S. J. Pearton, Gordon and Breach, New York, 1 (1997).

[5] J. H. Edgar, *Properties of group III nitrides*, edited by J.H.Edgar, IEE INSPEC, London (1994).

[6] R. Dingle, R. Shaklee, R. F. Leheny and R. B. Zetterstrom, *Appl. Phys. Lett.*, **19**, 5 (1971).

[7] H. Amano, I. Akasaki, T. Kozawa, K. Hiramatsu, N. Sawaki, K. Ikeda and Y. Ishii, *J. Lumin.*, **40-48**, 121 (1988).

[8] H. Amano, M. Kito, K. Hiramatsu and I. Akasaki, *Jpn. J. Appl. Phys.*, **28**, L2112 (1989).

[9] S. Nakamura, G. Fasol, *The Blue Laser Diode: GaN based light emitters and lasers*, Springer (1997).

[10] I. Grzegory, J. Jun, St. Krukowski, M. Bockowski and S. Porowski, *Physica B*, **185**, 99 (1993).

[11] I. Akasaki and H. Amano, *Tech. Dig. Int. Electron Devices Meet.*, **96**, 231 (1996).

[12] A. V. Blant, T. S. Cheng, C. T. Foxon, J. C. Bussey, S. V. Novikov and V. V. Tretyakov, Mater. Res. Soc. Symp. Proc., Symposium **449**, 465 (1997).

[13] H.P. Maruska and J.J. Tietjen, *Appl. Phys. Lett.*, **15**, 327 (1969).

[14] T. Detchprohm, K. Hiramatsu, K. Itoh and I. Akasaki, *Jpn. J. Appl. Phys.*, **31**, 1454 (1992).

[15] V.A. Savastenko and A.U. Sheleg, *Phys. Status Solidi A*, **48**, 135 (1978).

[16] A. Polian, M. Grimsditch and I. Grzegory, *J. Appl. Phys.*, **79**, 3343 (1996).

[17] Y. Takagi, M. Ahart, T. Azuhata T. Sota, K. Suzuki and S. Nakamura, *Physica B*, **219/220**, 547 (1996).

[18] A. F. Wright, *J. Appl. Phys.*, **82**, 2933 (1997).

[19] M. Yamaguchi, T. Yagi, T. Azuhata, T. Sota, K. Suzuki, S. Chichibu and S. Nakamura, *J. Phys. C*, **9**, 241 (1997).

[20] R. B. Schwarz, K. Khachaturyan and E. R. Weber, *Appl. Phys. Lett.*, **70**, 1122 (1997).

[21] C. Deger, E. Born, H. Angerer, O. Ambacher, M. Stutzman, J. Hornsteiner, E. Riha, and G. Fischerauer, *Appl. Phys. Lett.*, **72**, 2400 (1998).

[22] T. Deguchi, D. Ichiryu, K. Toshikawa, K. Sekiguchi, T. Sota, R. Matsuo, T. Azuhata, T. Yagi, S. Chichibu and S. Nakamura, *J. Appl. Phys.*, **86**, 1860 (1999).

[23] S. Stepanov, *MRS I.J.N.R.*, **6**, 1 (2001).

[24] D. Elwell and M.M Elwell, *J. Cryst. Growth*, **17**, 53 (1988).

[25] G. D. Chen, M. Smith, J. Y. Lin, H. X. Jiang, S. H. Wei, M. A. Khan and C. J. Sun, *Appl. Phys. Lett.*, **68**, 2784 (1996).

[26] S. H. Wei and A. Zunger, *Appl. Phys. Lett.*, **69**, 2719 (1996).

[27] P. Perlin et al., *Phys. Rev. Lett.*, **75**, 296 (1995).

[28] J. Neugebauer and C. G. Van de Walle, *Mater. Res. Soc. Symp. Proc.*, **339**, 687 (1994).

[29] S. Nakamura, T. Mukai and M. Senoh, *Jap. J. Appl. Phys.*, **31**, 2883 (1992).

[30] S. Nakamura, T. Mukai and M. Senoh, *Jap. J. Appl. Phys.*, **30**, 1708 (1991).

[31] T.L. Tansley and C.P. Foley, J. Appl. Phys., **59**, 3241 (1986).

[32] K. Kubota, Y. Kobayashi and K. Fujimoto, *J. Appl. Phys.*, **66**, 2984 (1989).

[33] S. Strite, D. Chandrasekhar, D.J. Smith, J. Sariel, H. Chen and N. Teraguchi, *J. Cryst. Growth*, **127**, 204 (1993).

[34] S. Adachi, *J. Appl. Phys.*, **58**, R1 (1985).

[35] A. U. Sheleg and V. A. Savastenko, *Inorg. Mat.*, **15**, 1257 (1979).

[36] R. R. Reeber and K. Wang, *MRS I.J.N.R.*, **6**, 3 (2001).

[37] H. J. Hovel and J. J. Coumo, *Appl. Phys. Lett.*, **20**, 71 (1972).

[38] J. W. Trainor and K. Rose, *J. Electron. Matter.*, **3**, 821 (1974).

[39] C. M. Wolfe, N. Holonyak and G.E. Stillman, *Physical Properties of Semiconductors*, Prentice Hall, Englewood Cliffs (1989).

[40] I. Akasaki and H. Amano, *GaN*, Academic, New York, **1**, 459 (1998).

[41] S. J. Pearton, F. Ren, J. C. Zolper and R. J. Shul, *J. Appl. Phys.*, **86**, 1 (1999).

[42] G. MArtin, A. Botchkarev, A. Rockett and H. Morkoc, *Appl. Phys. Lett.*, **68**, 2541 (1996).

[43] R. D. King-Smith and D. Vanderbilt, *Physical Review B*, **47**, 1651 (1990).

[44] R. Resta, *Rev. Mod. Phys.*, **66**, 899 (1994).

[45] F. Bernardini and V. Fiorentini, *Physical Review B*, **57**, 1 (1998).

[46] H. Morkoç, R. Cingolani, W. Lambrecht, B. Gil, H.-X Jiang, J. Lin, D. Pavlidis, K. Shenai, *MRS I.J.N.R.*, **4S1**, G1.2 (1999).

[47] M. Fanciulli, T. Lei and T. D. Moustakas, *Physical Review B*, **48**, 15144 (1993).

[48] W. Shan, B. D. Little, J. J. Song, Z. C. Feng, M. Schurmann and R. A. Stall, *Appl. Phys. Lett.*, **69**, 3315 (1996).

[49] M. D. McCluskey, C. G. Van de Walle, C. P. Master, L. T. Romano and N. M. Johnson, *Appl. Phys. Lett.*, **72**, 2725 (1998).

[50] T. Matsuoka, H. Tanaka, T. Sasaki and A. Katsui, *Inst. Phys. Conf. Ser.*, **106**, 141 (1990).

[51] M. Schuster, P. O. Gervais, B. Jobst, W. Hosler, R. Averbeck, H. Riechert, A. Iberl and R. Stommer, *J. Phys. D: Appl. Phys.*, **32**, A56 (1999).

[52] S. S. Lester, F. A. Ponce, M. G. Craford and D. A. Steigerwald, *Appl. Phys. Lett.*, **66**, 1249 (1995).

[53] C. Sasaoka, H. Sunakawa, A. Kimura, M. Nido and A. Sakai, *J. Cryst. Growth*, **189-190**, 61 (1998).

[54] S. C. Jain, K. Pinardi, H. E. Maes, R. Van Overstraeten, M. Willander and A. Atkinson, *Mater. Res. Soc. Symp. Proc.*, **482**, 875 (1998).

[55] S. Chichibu, T. Azuhata, T. Sota and S. Nakamura, *Appl. Phys. Lett.*, **70**, 2822 (1997).

[56] T. Takeuchi, C. Wetzel, S. Yamaguchi, H. Sakai, H. Amano and I. Akasaki, *Appl. Phys. Lett.*, **73**, 1691 (1998).

Chapter 6

InGaN/GaN Semiconductor Devices

6.1 Introduction

Progress in assessing the device potential of a wide band gap semiconductor material is impeded by the lack of detailed knowledge of the most important input parameters. As a consequence, simulations of nitride devices have to look at a range of values of some of the input parameters for which several values have appeared in the literature.

In order to validate and tune our model described in chapter 2, we have focused our attention on a single quantum well structure. The simplicity of this structure and the numerous literature on it make the single quantum well an ideal system for testing our software. Nevertheless, this simple system raises many issues for device simulation. An example is the problem of the value of the piezoelectric field. Strain, relaxation, elastic and piezoelectric constants, Poisson's ratio and bowing parameters are the ingredients of a complex recipe for evaluating the piezoelectric field. Using the single quantum well system will allow us to understand the role of each of these parameters in order to do a device simulation that can be at the same time in agreement with experimental results and predictive.

The chapter is divided in two parts. In the first part we present a detailed study of

a InGaN/GaN quantum well placed within a *pn*-junction. Section 6.2.1 introduces briefly the assumption of pseudomorphic growth of the InGaN layer while section 6.2.2 deals with the problem of the uncertainty of the value of the Poisson's ratio. Section 6.2.3 explains the effect of polarization on the band structure of the quantum well and the Quantum Confined Stark Effect (QCSE). In sections 6.2.4 and 6.2.5 we present our simulations of a 3 and a 5 nm quantum well, respectively. We interpret our results within the QCSE model even if screening mechanisms of the piezoelectric field have to be considered in order to obtain agreement with experimental results. The magnitude of the screening is shown to vary with the well width. A closer comparison of quantum wells with different widths is done in section 6.2.6. Section 6.2.8 introduces the problem of indium fluctuations in a quantum well and their effect on the optical band gap. Misfit dislocations are briefly mentioned in the section 6.2.9.

In the second part of this chapter (section 6.3) an extended simulation of a novel InGaN/GaN double-well LED design has been carried out. This device, using asymmetric tunneling of electrons and holes, promises higher performance than common GaN-based LEDs. After a description of the functionality of the device, section 6.3.1 presents our simulations. Particular care has been given to the role of the piezoelectric field on the band profile of the device. Resonance conditions are found to be highly sensitive to small variations of piezoelectric field. Furthermore, build-up of strain-induced charge prevents optimal electron injection. Different designs have been tested. Section 6.3.2 is aimed to understand the effect of doping of the active region of the CART device.

We conclude the chapter with a analysis of the effect of traps on the current-voltage characteristic of GaN *pn*-junctions.

6.2 Single Quantum Well Structure

The single quantum well device that will be the object of all our simulations consists of a *n*-GaN/qw-$In_xGa_{1-x}N$/*p*-GaN structure with 0.5 μm thickness for the 10^{18} cm^{-3} *n*- and *p*-doped regions. The quantum well has been lightly *n*-doped (10^{16} cm^{-3}) and its thickness and indium content have been varied.

6.2.1 Relaxation

Using the theoretical model of Fischer et al.[1], relaxation of the biaxially compressed layer should occur for critical layer thicknesses in the range 4-14 nm (x=0.1) and 6.3 nm (x=0.2) when room temperature stiffness constants of GaN are assumed. However, experimentally it has been shown that no relaxation is observed for 60-70 nm thick InGaN layers for $x < 0.4$ [2][3][4]. All the lattice parameters for InGaN are close to their GaN equivalents, indicating that the InGaN layers can be grown pseudomorphically on the underlying GaN buffer layer. We are not aware of any model that could explain the large thickness of the unrelaxed layers and tentatively assume this is due to the higher degree of ionicity contributing to the bonding forces in the nitrides and the uncertainties of the stiffness constants at the growth temperature. For these reasons, pseudomorphic growth of InGaN has been considered.

6.2.2 Poisson's ratio and Polarization

While experimentally the value of the indium content x is derived by the measurement of the lattice constant c and a (see chapter 5 section 5.4.4), as far as simulation is concerned x is an input of the program. Therefore, in the hypothesis of pseudomorphic growth of InGaN on GaN, knowing that $a = a(GaN)$, applying the equation 5.10 in chapter 5 we get the value of c. Using c, a and x we determine the value of the piezoelectric polarization P_{pz} as demonstrated in chapter 2 section 2.9. The piezoelectric polarization is then given by

$$P_{pz} = e_{33}(x)\left(\frac{c}{c(x)} - 1\right) + 2e_{31}(x)\left(\frac{a}{a(x)} - 1\right) \tag{6.1}$$

where the in-plane lattice constant $a(x)$, the c-axis lattice constant $c(x)$, the piezoelectric constants $e_{33}(x)$ and $e_{31}(x)$ are all evaluated as linear interpolation between the respective GaN and InN values (Vegard's Law) (see tables 5.2.7 and 5.3.3 in chapter 5 for the values of the parameters used).

As reported in table 5.2.3 of chapter 5, there is a large discrepancy in the literature

on the value of the Poisson's ratio. We have run a few simulations in order to have a qualitative picture of the influence of the Poisson's ratio on the piezoelectric polarization.

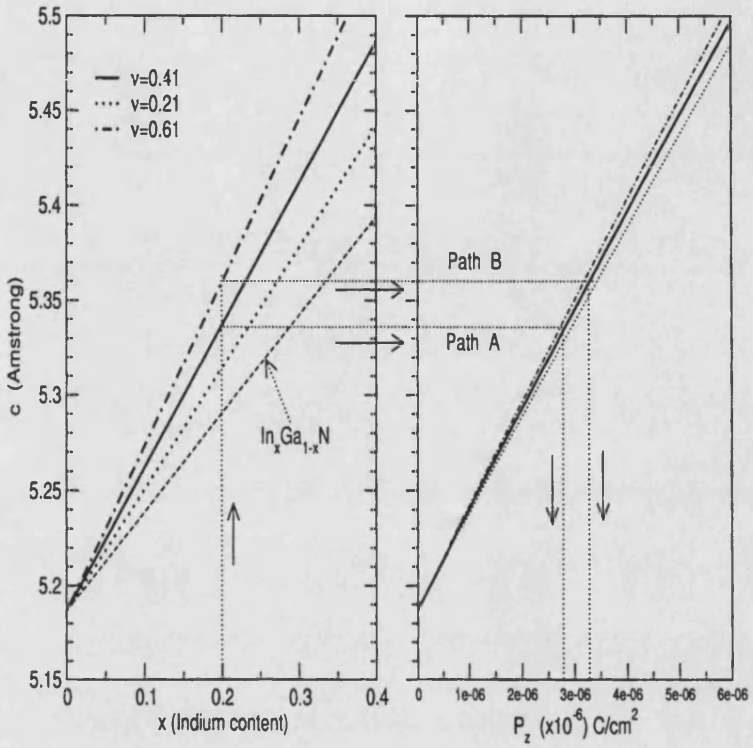

Figure 6.1: (Left panel) c-axis lattice constant versus indium content for three values of the Poisson's ratio. The dashed line represents the relaxed value of c. (Right panel) c-axis lattice constant versus the piezoelectric polarization for three values of the Poisson's ratio: 0.21, 0.41, 0.61

Figure 6.1 (left panel) shows the c constant versus the indium content x. In deriving c from a and x a value for the Poisson's ratio ν has to be assumed. We have used ν=0.41 as the reference value (solid line in the left panel of figure 6.1) as suggested by Stepanov et al.[5]. Two other lines are plotted in figure 6.1(left), respectively for ν=0.21 (dotted line) and ν=0.61 (dot-dashed line). These values have been freely chosen to demonstrate the dependence of c on the choice of ν. Calculations show that, even if the value of c is strongly affected by a variation of ν (left panel of figure 6.1), the piezoelectric polarization does not follow the same pattern. This can be seen in figure 6.1 (right panel) where c is plotted versus P_{pz}. In this graph the two lines for ν=0.21 and ν=0.61 shows little difference from the case of ν=0.41 (solid central line). However, the uncertainty on c can lead to noticeably different values of piezoelectric polarization as illustrated by the

two arrow-paths in figure 6.1. If we fix x=0.2 and we use two different values of Poisson's ratio, respectively ν=0.41 in path A and ν=0.61 in path B of figure 6.1, the variation on c is transmitted to the final value of piezoelectric polarization.

In order to estimate the contribution of the in-plain a strain and the perpendicular c strain the analysis of figure 6.2 is of help.

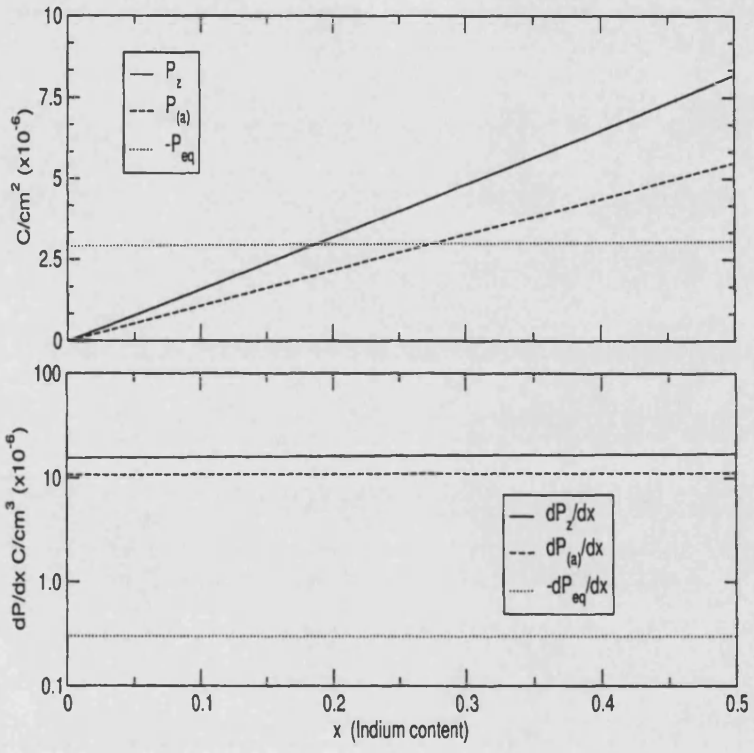

Figure 6.2: (Top Panel) Piezoelectric polarization P_{pz}, spontaneous polarization P_{eq} and part of the piezoelectric polarization derived from the in-plane strain $P_{(a)}$, versus indium content. (Bottom Panel):Derivatives of P_{pz},P_{eq} and $P_{(a)}$.

In equation 6.1 it is possible to identify two terms: the term dependent on a that we call $P_{(a)}$ and the term $P_{(c)}$ function of c-axis lattice constant,

$$P_{pz} = P_{(c)} + P_{(a)} \tag{6.2}$$

where

$$P_{(a)} = 2e_{31}(x)\left(\frac{a}{a(x)} - 1\right) \tag{6.3}$$

$$P_{(c)} = e_{33}(x)\left(\frac{c}{c(x)} - 1\right). \tag{6.4}$$

Figure 6.2 (top panel) plots P_{pz} and $P_{(a)}$ as function of x. Graphically, $P_{(c)}$ is the difference between P_{pz} and $P_{(a)}$. It easy to spot that $P_{(a)}$ is the main contribution to the total piezoelectric polarization.

The dotted quasi-horizontal line in figure 6.2 (top panel) represents the inverse of the equilibrium polarization P_{eq}. The equilibrium polarization, otherwise known as the *spontaneous polarization*, has the opposite sign compared to the strain induced polarization for InGaN/GaN systems grown along the <0001> axis. It would at first sight appear that P_{eq} is comparable with piezoelectric polarization for a large range of indium content, therefore spontaneous polarization should play an important role in determining the field distribution of the device structure. This is not totally correct for the reason that the piezoelectric and spontaneous polarisations enter the drift-diffusion equations only in the form of their gradient as it can be seen in the Poisson equation (equation 2.5 of chapter 2). A polarization field is induced by a charge dipole caused by the displacement of atoms in the lattice. This displacement can be either the result of bonding forces within the lattice (spontaneous polarization) or due to an externally applied stress (piezoelectric polarization). At the heterointerface between two materials with different polarization (i.e different dipole charge), a fixed charge proportional to the first derivative of the polarization will form. The derivatives of P_{pz}, $P_{(a)}$ and P_{eq} are plotted in figure 6.2 (bottom panel). $\mathrm{d}P_{eq}/\mathrm{dx}$ is almost two orders of magnitude less than $\mathrm{d}P_{pz}/\mathrm{dx}$ and compensation of P_{pz} can not come from P_{eq}. For this reason many authors neglect P_{eq} as a first approximation [3]. Nevertheless, our model includes spontaneous polarization.

6.2.3 The value of Piezoelectric Polarization

The piezoelectric polarization P_{pz} has a large impact on the band profile of a single quantum well structure. This can be seen in three different examples described schematically in figure 6.3. Figure 6.3(a) shows the conduction and valence band diagram for a quantum well within a *pn*-junction when no strain-induced electric

field E_{pz} is present. In this case the bending of the band is due to the built-in electric field E_i of the junction. Consequently the wavefunctions of the ground states C_1 in the conduction band and V_1 in the valence band, move apart, away from the centre of the quantum well. The overlap of the wavefunctions is therefore decreased. Furthermore, as the wavefunctions move apart, the energy levels shift down towards the bottom of the quantum well. This is the well known *Quantum Confined Stark Effect* (QCSE). Experimentally this is seen as a red-shift in the emission or absorption spectrum for the effect of a diminished optical bandgap (C_1-V_1) [6]. Consequently, a decrease of the emission intensity due to the lower overlap and oscillator strength is observed too.

As a forward bias is applied to the *pn*-junction, the induced electric field $E_{bias} = -dV_{bias}$ (d is the well thickness and V_{bias} the applied voltage) opposes the built-in electric field. Eventually, the flat band condition is reached when the applied voltage matches the built-in potential.

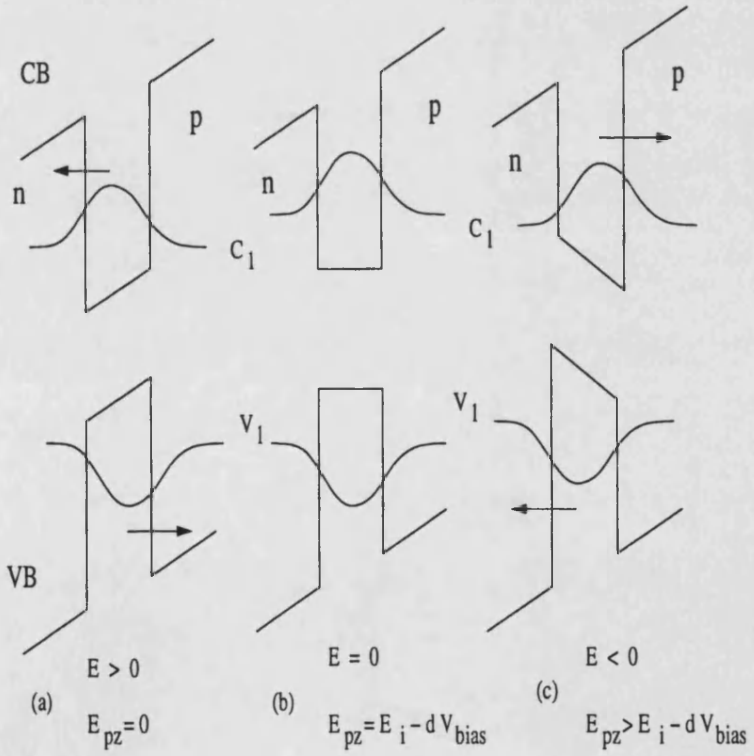

Figure 6.3: Conduction and valence band profile for a single quantum well for (a) $E_{pz} = 0$, (b) $E_{pz} = E_i + E_{bias}$, (c) $E_{pz} > E_i + E_{bias}$

However, flat band conditions can be obtained for any bias in the presence of a piezoelectric field as shown in figure 6.3(b). It is possible to demonstrate that for an InGaN quantum well growth on a <0001> GaN substrate, the direction

of the strain-induced piezoelectric field is from the top InGaN layer to the GaN substrate [6]. In our system of a quantum well embedded in a *pn*-junction, the piezoelectric field E_{pz} is directed from the *p*- to the *n*-type region. The flat band condition then becomes

$$E_{pz} - E_i + E_{bias} = 0. \tag{6.5}$$

Figure 6.3(c) describes the band profile when the piezoelectric field is much larger than the other two fields. As in figure 6.3(c), this leads to QCSE and a decrease of the wavefunction overlap.

We have simulated a single $In_xGa_{1-x}N$ quantum well structure within a 1 μm n-doped ($10^{18}cm^{-3}$) GaN layer. No bias was applied. Parameters of the materials were reported in the previous chapter. The solver used was the one-dimensional drift diffusion model with Schrodinger equation solved self-consistently. Figure 6.4 shows the magnitude of the red-shift due to the QCSE for two quantum wells with different indium content x as a function of the well width. The red-shift is given by the difference between the optical band gap for a quantum well in the presence (strained line) and in the absence (relaxed line) of the polarization. Since for these calculations we do not consider any built-in field, the case of a well fully relaxed can be represented by the square quantum well of figure 6.3(b) while the case of the strained quantum well is depicted in figure 6.3(c).

As can be seen from figure 6.4, the QCSE-induced red shift depends strongly on the well width and on x. For x=0.1 (top two lines of figure 6.4), the red-shift increases linearly with the well width and tends to a constant value as the C_1 and V_1 levels approach the well bottom. For x=0.2 (bottom two lines of figure 6.4), the red-shift variation has a steeper gradient than the previous case and no saturation can be seen. This result is in agreement with experiments [7] and can be explained with the passage from a square quantum well to a triangular quantum well.

A quick calculation provides an order of magnitude of piezoelectric field in a quantum well. Using Bernardini's parameters [8] of tables 5.2.7 and 5.3.3 in chapter 5, in a $In_{0.2}Ga_{0.8}N$ quantum well and assuming pseudomorphic growth, we obtain a piezoelectric polarization of $P_{pz,Max}$=2.75$\times$ 10^{-6} C/cm^2. Consequently,

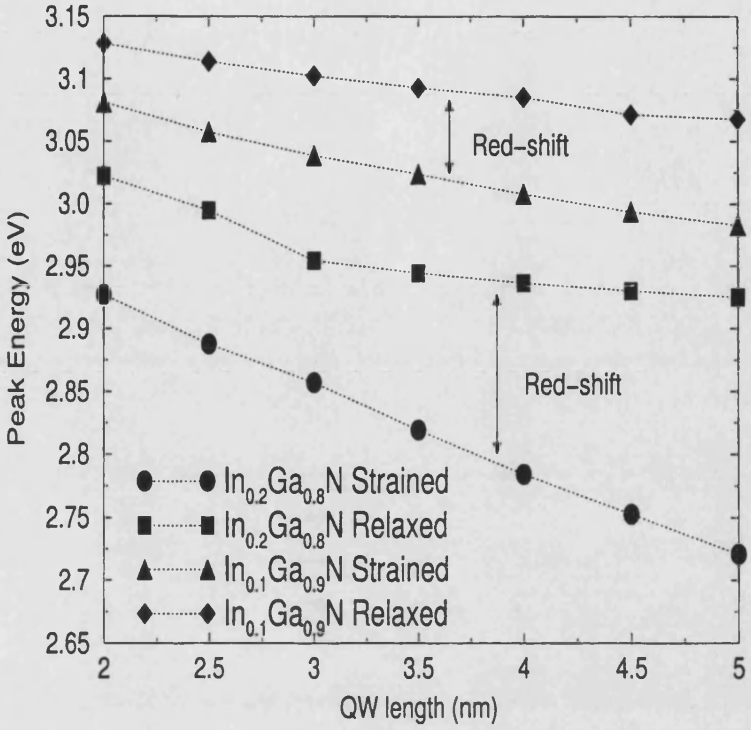

Figure 6.4: Optical gap versus well width for $In_{0.1}Ga_{0.9}N$ and $In_{0.2}Ga_{0.8}N$ relaxed/strained quantum wells.

the piezoelectric field E_{pz} is given by

$$E_{pz,Max} = \frac{P_{pz,Max}}{\varepsilon_0 \varepsilon_r} = 3.05 \quad \text{MV/cm} \tag{6.6}$$

where ε_r is the relative dielectric constant. The subscript Max will refer to the maximum value of piezoelectric field or piezoelectric polarization that will be used in our calculations. More precisely, since the values of the piezoelectric polarization and field depend on the indium content of the quantum well, we will take as our maximum value the result of the calculation of P_{pz} using equation 6.1 where ε_{31} and ε_{33} are taken from tables 5.2.7 and 5.3.3 of chapter 5. In deriving the piezoelectric field, equation 6.6 will be used. As far as the equilibrium polarization is concerned, we define $P_{eq,Max}$ using those values for GaN and InN (or their linear combination for InGaN) as reported by Bernardini et al.[8] and summarised in the previously cited tables. Consequently, the maximum total polarization would be $P_{Max} = P_{pz,Max} + P_{eq,Max}$.

The value of E_{pz} reported in equation 6.6 is extremely high if compared to the

values of E_i and E_{bias} which are usually in the range of few hundreds of kV/cm. This being the case, figure 6.3(c) would be a close approximation of the quantum well band profile. A very high QCSE and a low wavefunction overlap are expected.

A higher doping concentration in the bulk p- or n-type regions can help increase both the optical bandgap and the overlap. This is due to the fact that an effect of high doping concentrations is the reduction of the depletion region width. The built-in potential is almost constant and a higher E_i is necessary to produce the same difference of potential in the shorter depletion region. From Poisson equation (2.5), the fixed dopant charge varies as the gradient of the electric field. Therefore, an increased doping concentration leads to a larger E_i which is more effective in screening the piezoelectric field and reducing the QCSE.

6.2.4 3 nm Quantum Well

Figure 6.5 shows the optical bandgap (or the wavelength) of the transition between the first level in conduction band C_1 and the first level in valence band V_1 versus the fraction of polarization for a 3 nm quantum well. There are three groups of three curves. Each group represents a different percentage of indium content x in the quantum well. In each group, each line corresponds to a different bias. The horizontal axis is in units of P_{Max}. Note that the maximum value of polarization depends on x so curves representing different values of x in the quantum well will not share the same scale of polarization intensity. Poisson's ratio has been set to 0.205.

An important observation is that, for a given indium content, there is a maximum transition energy (or equivalently a minimum wavelength). This maximum corresponds to the case of the flat band condition shown in figure 6.3(b) and defined in equation 6.5.

For the sections of the curves plotted in figure 6.5 that lie on the left-hand side of the maxima, the band profile of the quantum well resembles the schematic diagram of figure 6.3(a). Here, E_i prevails over E_{bias} and E_{pz}. The right hand side, negative slope of the curves of 6.5 represents the case where the piezoelectric field is dominant. For higher biases, E_{bias} increases and a lower field, i.e. a lower

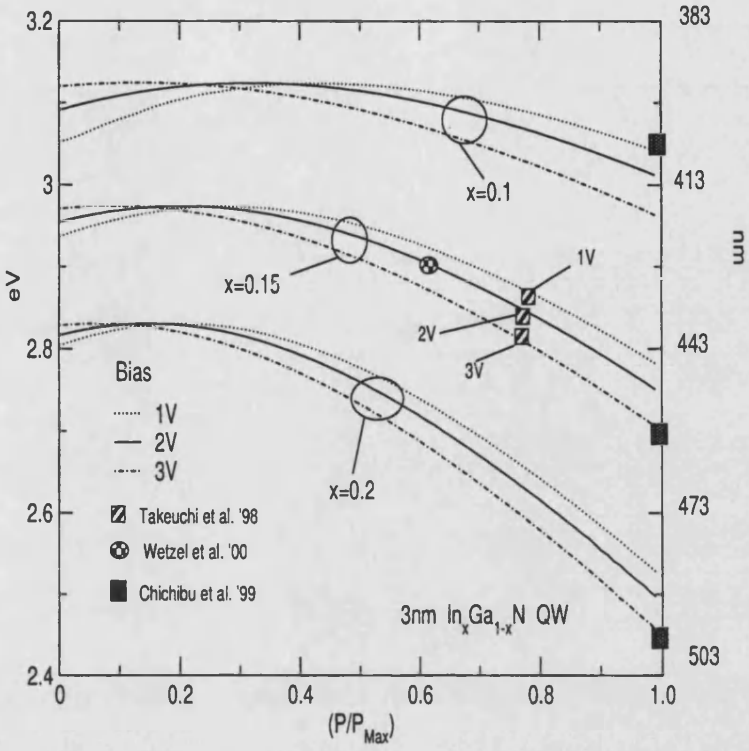

Figure 6.5: Optical bandgap(C_1-V_1 transition) for 3 nm $In_xGa_{1-x}N$ quantum well as a function of the fraction of polarization as in Bernardini [8]. x=0.1, x=0.15, x=0.2 at 1 V (dotted lines) , 2 V (solid lines) and 3 V (dashed-dotted lines) bias. Boxes represent experimental results from Takeuchi et al.[6]

polarization, is necessary to achieve the flat band condition as shown in figure 6.5 or as derived directly from equation 6.5. The decrease of the optical band gap due to the QCSE induced by high piezoelectric fields is more pronounced in deeper wells obtained by an high indium concentration x. From figure 6.5 it is possible to estimate that for the case of a quantum well with indium content of x=0.1, the decrease of optical band gap from the flat band condition to the full value of polarization applied is less than 0.15 eV. For the case of a quantum well with x=0.2, the optical band gap variation can be as large as 0.35 eV.

Overlap calculations are shown in figure 6.6 corresponding to the same data as in figure 6.5. As expected, in the flat-band condition the overlap is maximised. A modest increment of overlap is found when the indium composition increases as a result of the enhanced confinement effect in the well.

Figure 6.7 plots the electric field for the 3 nm quantum well. This time each panel corresponds to a fixed bias whereas the three lines plotted are derived using a different indium content. In accordance with equation 6.5, for P/P_{Max}=0

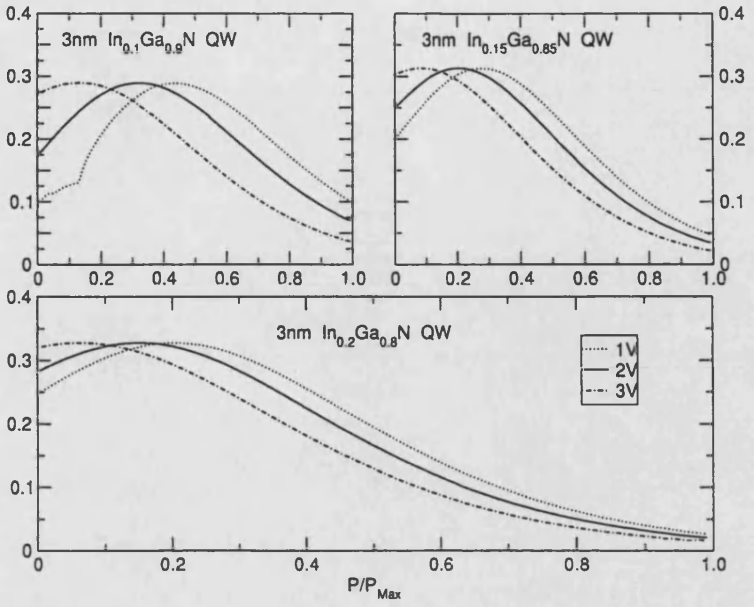

Figure 6.6: Wavefunction overlap for the C_1-V_1 transition, as a function of the fraction of maximum polarization for 3nm $In_xGa_{1-x}N$ quantum well. In the top left panel x=0.1; in the top right x=0.15. Bottom panel has x=0.2. In each panel the three curves represent different bias: 1 V (dotted line), 2 V (solid line) and 3 V (dot-dashed line).

the piezoelectric field is zero and the value reported (positive direction from the substrate to the top layers) is the sum of the built-in field minus the field deriving from the forward bias applied. Values of electric field over 2.5 MV/cm are obtained for x=0.2. Such a high field can not be screened by any applied voltage.

One of the aims of the single quantum well simulation was to be able to find the value of the piezoelectric field from information about the emission or absorption spectra experimentally obtained. Figure 6.5 can be used for this purpose.

The analysis of experimental results of Takeuchi et al.[6] reveals many issues regarding the value of the piezoelectric field inside a quantum well. They performed a series of measurements on a 3 nm $In_{0.15}Ga_{0.85}N$ quantum well under different applied biases. Results for 1, 2 and 3 Volts of applied bias are plotted in figure 6.5. They observed blue-shifts of the PL peak energies in a 3 nm $In_{0.15}Ga_{0.85}N$ quantum well, grown either on AlN buffer layer and a GaN substrate by hydride vapour phase epitaxy (HVPE), with increasing applied reverse voltage. In accordance with our simulation, this phenomenon can be explained by the fact that the applied reverse voltage cancels the internal piezoelectric field. The reverse bias produces an electric field that has the same sign as the built-in field and

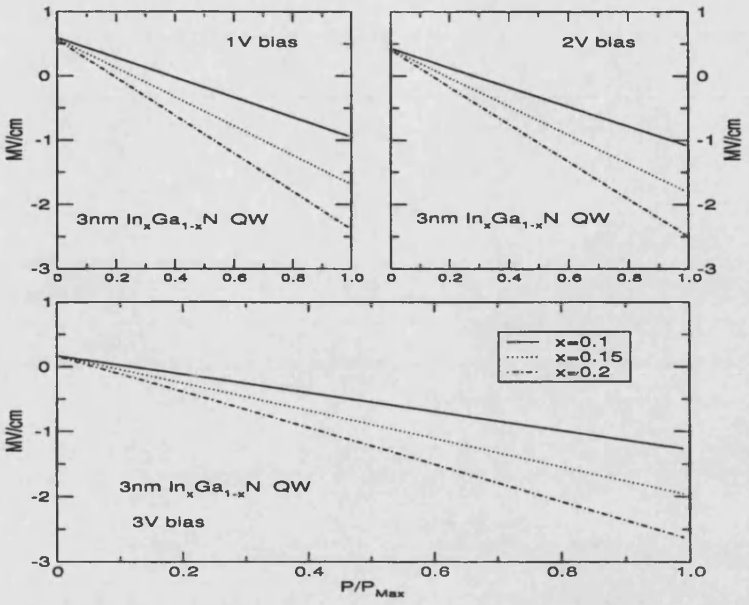

Figure 6.7: Electric field in 3 nm $In_xGa_{1-x}N$ quantum well, as a function of the fraction of the maximum polarization and applied bias. In each panel the three lines represent different indium content: x=0.1 (solid line), x=0.15 (dotted line) and x=0.2 (dot-dashed line).

it is opposite to the piezoelectric field. We have not run any simulations with reverse bias, but the trend should be clear by analysing figure 6.5 when the bias is decrease from 3 V to 1 V.

The data reported by Takeuchi et al.[6] agrees well with our simulations for a value of polarization that is around 75-80% of P_{Max} for a $In_{0.15}Ga_{0.85}N$ quantum well. In the same work [6], a calculated value of piezoelectric field was obtained of about -1.2 MV/cm from reverse bias measurements and between -1.4 MV/cm and -1.6 MV/cm for forward bias. If we compare these values to our calculated electric field in figure 6.7, and if we subtract the built-in field (that is the field for zero polarization) from the curves of figure 6.7, again we find a value for the total polarization around 75%. Reference [6] confirms that there is a discrepancy between the expected theoretical piezoelectric field (in the range 2.7-2.9 MV/cm) and the experimentally obtained value. It is argued in [6] that this discrepancy could be attributed to the screening of the piezoelectric fields by free and photo-generated carriers.

Wetzel et al.[9] give a transition of 2.9 eV in a 3 nm $In_{0.15}Ga_{0.85}N$ quantum well with a field of 0.6 MV/cm. From figure 6.5 we deduce a value of 0.6 for P/P_{Max} at a 2.9 eV transition when x=0.15 and the bias is around 1 V or less. In figure 6.7, for P/P_{Max}=0.6 we find that the total field is around 0.65 MV/cm in reasonable

agreement with the experimental data. Of course the piezoelectric field is higher if we consider the built-in field.

The fundamental transition energies in a single 3 nm quantum well reported by Chichibu et al.[7], 2.4 eV, 2.7 eV and 3.05 eV respectively for x=0.22, x=0.15 and x=0.1, correspond in our calculations to the maximum polarization $P = P_{Max}$. It is interesting to observe that the values of piezoelectric fields they obtained do not exceed 0.4 MV/cm, far below the 3 MV/cm of Bernardini's calculations [8]. The authors explain this discrepancy arguing the existence of an internal screening mechanism coming from poorly identified internal charges [7]. Reference [7] addresses the problem of reduction of the oscillator strength due to the fact that electron and holes are confined to opposite sides of the well.

6.2.5 5 nm Quantum Well

Using excitation power spectroscopy Perlin et al.[10] obtained emission between 2.325 eV and 2.45 eV for a 5.3 nm $In_{0.15}Ga_{0.85}N$ quantum well. For a $In_{0.1}Ga_{0.9}N$ 4.5 nm the emission was around 2.80 eV and 2.85 eV. The variation of the value was attributed to the screening effect due to the optical excitation of the quantum well.

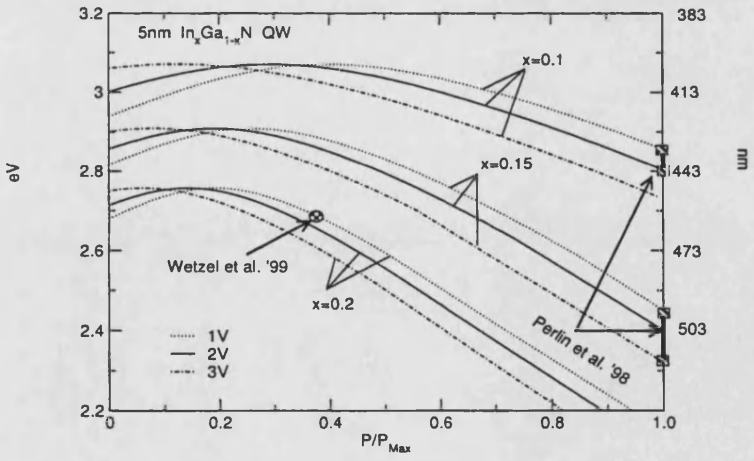

Figure 6.8: Optical bandgap(C_1-V_1 transition) for 5 nm $In_xGa_{1-x}N$ quantum well as a function of the total polarization in units of P_{Max}. Indium content is varied as x=0.1, x=0.15, x=0.2 at 1 V (dotted lines), 2 V (solid lines) and 3 V (dashed-dotted lines) bias.

This screening mechanism has been reported also by others [11] and consists of the build up of charge in the quantum well due to electron-hole generation by optical excitation. This excess of charge gives rise to an electric field that opposes the piezoelectric field and reduce the QCSE. We have compared Perlin's results with our 5 nm quantum well simulation. The data fit our simulated curve for the maximum value of polarization ($P = P_{Max}$) as shown in figure 6.8.

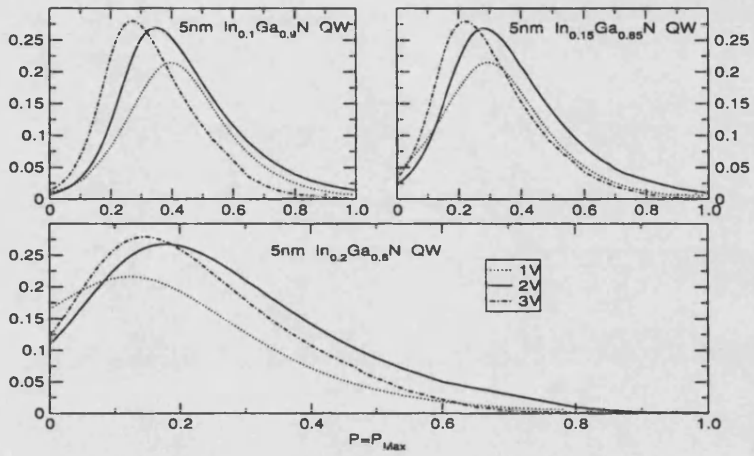

Figure 6.9: Wavefunction overlap for the C_1-V_1 transition, as a function of the total polarization in units of P_{Max} for 5 nm $In_xGa_{1-x}N$ quantum well. In the top left panel x=0.1; in the top right x=0.15. The bottom panel has x=0.2. In each panel the three curves represent different biases: 1 V (dotted line), 2 V (solid line) and 3 V (dot-dashed line).

Figure 6.9 plots the overlap calculations for the 5 nm quantum well as in figure 6.8. An interesting result is that for the maximum value of polarization, the overlap is less than 0.02. This would lead to a low intensity emission. Intensity measurements published by Perlin et al.[10] were taken at low temperature (T$\sim$4 K). They report a linear dependence between peak intensity and excitation power. We believe that the low value of overlap may explain the necessity of a low-temperature measurement and the strong peak intensity and excitation power correlation obtained.

Calculations of electric field in the 5nm quantum well agree with the results of Perlin et al.[10] showing a high piezoelectric field over 1 MV/cm as reported in figure 6.10. Wetzel et al.[9][12] found C_1-V_1=2.63 eV from PR and PL measurements of a 5nm InGaN quantum well with x=0.2 (see figure 6.8). They calculated that this optical bandgap is associated to a piezoelectric field of 0.9 MV/cm [9]. Checking these values against the simulation of figures 6.8 and 6.10, P/P_{Max}=0.4

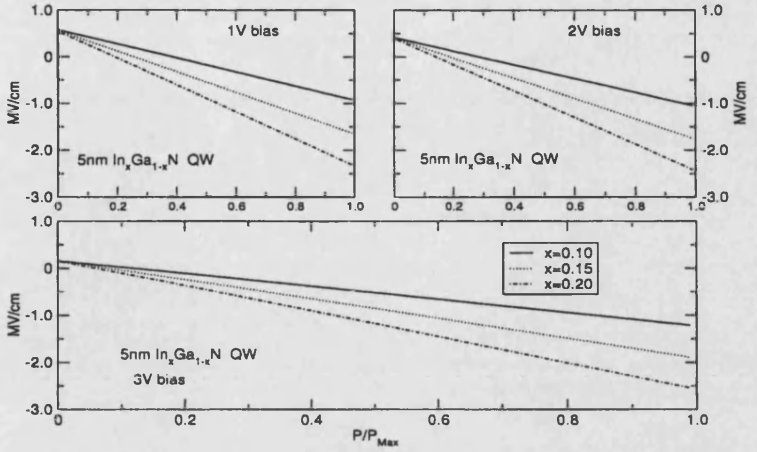

Figure 6.10: Electric field in 5 nm $In_xGa_{1-x}N$ quantum well, as a function of P/P_{Max}. 1 V bias in the top left panel ; 2 V bias in the top right x=0.15. Bottom panel has 3 V bias curves. In each panel the three lines represent different indium content: x=0.1 (solid line), x=0.15 (dotted line) and x=0.2 (dot-dashed line).

fits results both for the optical band gap and for the electric field.

6.2.6 Optical band gap for different well widths

Previous calculations were done on a quantum well of a fixed width varying the indium content and the applied bias. In this section we directly compare results for three different well widths (5, 10 and 15 nm) for a fixed applied bias of 2 V and indium content x=0.2. As before, the quantum wells are lightly n-doped (10^{15} cm^{-3}) and are placed within a pn-GaN 1 μm junction doped at (10^{17} cm^{-3}) for both p and n sides. Figure 6.11 shows the variation of the optical bandgap with the polarization for these three quantum wells. The dotted line in figure 6.11 represents the optical band gap for a well width of 5 nm at a bias of 3 V. The inset to figure 6.11 shows the corresponding overlap values.

Figure 6.11 shows that for each well width there is a P at which we have both a maximum optical bandgap energy and maximum overlap. This corresponds to the flat band condition. The well width hardly alters the value of P at which this occurs, since neither the built-in field nor the piezoelectric field depend strongly on the well width. However the wider the well, the larger the variation of the optical band gap with P. The overlap is strongly dependent on P, with wider quantum wells having a lower overlap as can be seen in the inset of figure 6.11.

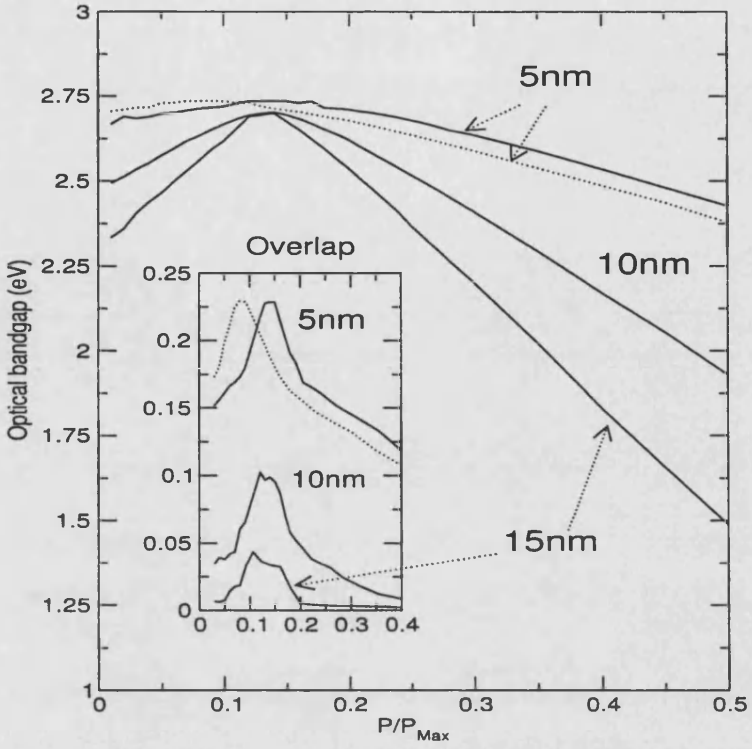

Figure 6.11: Optical bandgap(C_1-V_1 transition) as a function of the total polarization P for 5 nm, 10 nm, 15 nm GaN/$In_{0.2}Ga_{0.8}N$/GaN quantum well at 2 V bias. The dotted line is for a 5 nm quantum well under 3 V bias. The inset shows the respective overlap curves.

For a 15 nm thick quantum well made of $In_{0.2}Ga_{0.8}N$, Wang et al.[5] give C_1-V_1=2.7 eV. This result, compared with the simulations in figure 6.11, shows that in the sample there must have been almost full compensation of the piezoelectric field (for P/P_{Max} ~0.15). Unfortunately no measurement of the piezoelectric field was reported.

6.2.7 The quasi-cubic approximation

So far, we have shown that it is possible to deduce the value of the piezoelectric field for a particular quantum well by comparing the PL or PR experimental data with our simulations. We are able to estimate the magnitude of the screening involved and to obtain useful informations on the overlap of the wavefunctions. We have shown that different values of polarization and piezoelectric field can be deduced from several results found in the literature. This indicates that the screening mechanism is not constant and it strongly depends on the well width. We have shown also that the uncertainty in Poisson's ratio is not critical but

affects the value of piezoelectric polarization. Our next obvious step is to see whether the piezoelectric constants that we have assumed in our calculation (see tables 5.2.7 and 5.3.3 in chapter 5) are effectively correct. Moreover, if Bernardini's values are not correct, we need to know how changing them affects our results. This last problem has been already solved. In fact a variation in the piezoelectric constants leads to the same variation of the piezoelectric polarization and P. Almost all the results in the previous sections have been reported as a function of P. Consequently they can be interpreted thinking that the the variation in the polarization is due to the uncertainty in the piezoelectric constants and not to screening effects. The remaining question is whether it is safe to rely on Bernardini's calculation. Many authors do [11], others are more critical [13][14].

In order to compare piezoelectric properties of different semiconductors, it is common to use the cubic piezoelectric constant e_{14}. Rigorously, this coefficient describes the piezoelectric properties of a cubic lattice (eg. zincblende), but it is an acceptable approximation to use this parameter also for hexagonal lattices (eg. wurtzite). The *quasi-cubic approximation* is described by the following

$$e_{31} = -e_{14}/\sqrt{3}, \tag{6.7}$$

$$e_{33} = 2e_{14}/\sqrt{3}. \tag{6.8}$$

One of the first piezoresistive measurements made by Bykhovski et al.[15] reported e_{14}=0.56 C/m^2, a value noticeably lower than e_{14}=0.79 C/m^2 derived from the first principles calculations of Bernardini et al.[8]. Wetzel et al.[3] performed PL and PR measurements on a 40nm-$In_{0.187}Ga_{0.813}N$/GaN single heterostructure. Measuring the period of Franz-Keldysh oscillations in photoreflection they obtain piezoelectric field up to a maximum of 1.12 MV/cm. With these values they calculated e_{14}=0.1 C/m^2, in line with the values reported for other III-V and II-VI semiconductors. These results may either imply that Bernardini's piezoelectric constants are overestimated or that experimental measurements underestimate e_{14} because of the inter-play of screening mechanisms.

6.2.8 Indium fluctuations

Shapiro et al.[16] present PL measurements of a 2.5 nm quantum well with x=0.2. Emission from this quantum well was found at a low value of 2.28 eV and no blue-shift with excitation power was registered. The low value of energy emission does not fit our simulations. For a 3 nm thick quantum well with x=0.2 the optical band gap should not be lower than 2.4 eV as shown in figure 6.5. The absence of excitation power blue-shift raises concern too. We recall that a consequence of the presence of an electric field in a quantum well is that the thicker the well is, the greater the spatial charge separation, the potential drop across the well and the QCSE. This can be seen clearly when we compare figures 6.5 and 6.8. For the 5 nm quantum well the curves representing the optical bandgap in figure 6.8 have a steeper variation with the polarization than those for the 3 nm quantum well of figure 6.5. Even clearer proof of that is shown in figure 6.11. The variation of optical bandgap is an index of the magnitude of the QCSE. Therefore, the excitation induced blue-shift is expected to be stronger in structures with thicker wells where the high QCSE should be more sensitive to any kind of screening mechanism. Rough calculations show that a 2.5 nm thick quantum well with x=0.2 should still exhibit a piezoelectric-field induced potential drop of about 70% of that one found in a 5 nm broad well with x=0.15.

The absence of excitation power blue-shift in the previously cited experiments of Perlin suggests that QCSE is not the main mechanism responsible for the low level of PL emission [16]. It has been thought that a quantum well with large spatial indium fluctuations will not have a smooth, sloping band structure that is necessary for the QCSE. The carriers will reside in local minima created by the fluctuations instead of separating to opposite sides of the well. Thus, the QCSE will be reduced or even eliminated in wells with large indium fluctuations. This idea may explain also the low energy emission from the 2.5 nm well. If the emission is from a deep minimum in the well, the effective indium concentration is higher than the nominal x=0.2.

In order to investigate the effect of local variations of indium concentration x in a quantum well, we have simulated both random and periodic indium fluctuations up to 30% of the average value x=0.15 (i.e. x=0.15$\pm$0.05). The variation of the indium content was considered only in the growth direction, therefore no quantum dot-like structures were obtained. The transition energy and electric

field are reported in figure 6.12 for the 3 nm quantum well with 2 V applied bias. The electric field shown is the average value in the well. The length scale of the fluctuations has been set to 0.6 nm, that is roughly the value of the c lattice constant.

For a randomly distributed indium concentration the C_1-V_1 transition energy increases in comparison with the case without fluctuations. We explain this from the fact that the energy level in the quantum well tends to move away from the band bottom. From the overlap results we find that the wavefunction becomes more central in the well with a slight increase in overlap. We have seen that the blue-shift increases with the magnitude of the fluctuation. A modest increase of the electric field is obtained, due to the fact that locally the strain can assume higher values than when x is constant throughout the quantum well. No major changes are found for the sine-function pattern of indium content when the period of the oscillation T_{fluct} is much lower than the well width.

More interesting situations arise instead when T_{fluct} is comparable with the quantum well width. For a $\sin(x)$-like fluctuation a minimum in the well is formed near the interface with the n-doped GaN layer. The spatial extension of the minimum of the well is not large enough to allow localisation of the electron wavefunction. However, since holes have a higher effective mass, hole wavefunction localisation is possible and is helped by the direction of the piezoelectric field that confines holes near the n-GaN/InGaN heterointerface. Consequently, a strong red-shift in the C_1-V_1 transition is obtained as shown in figure 6.12. For a $-\sin(x)$-like fluctuation, the minimum forms far from the n-GaN/InGaN heterointerface and no hole wavefunction localisation is achieved. On the contrary, both electron and hole wavefunctions are pushed away from the well bottom and a blue-shift is obtained as shown in figure 6.12.

Whether the initial concentration of indium in the quantum well, (the one at the heterointerface with the n-doped GaN) first rises then decreases or vice-versa affects considerably the magnitude of the electric field. As seen in figure 6.12, fluctuations starting with an increase of indium content ($\sin(x)$) give rise to a higher piezoelectric field. The strain at the contact with the GaN n-layer is higher and so is the polarization. On the other hand, an initially low indium concentration ($-\sin(x)$) reduces the strain and the piezoelectric field. These results might explain why experiments measure high PL emission energies but

low levels of the piezoelectric field.

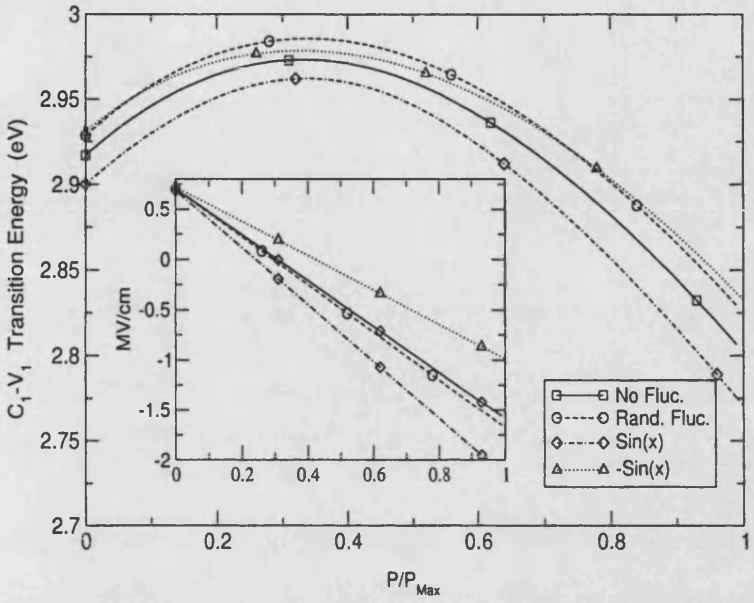

Figure 6.12: Top panel: Peak energy of the C_1-V_1 transition for a 3 nm $In_{0.2}Ga_{0.8}N$ quantum well at 2 V bias with and without fluctuations of indium content as function of the fraction of the maximum polarization. Random and sin-like fluctuations are considered. Bottom-panel: Electric field.

There are further issues that have to be addressed. Wetzel et al.[3], in order to identify the role of possible fluctuations and their contribution to the emission process, have performed a spatially resolved PL mapping at room temperature on a 40 nm-$In_{0.187}Ga_{0.813}N$/GaN single heterostructure. A spatial fluctuation of the peak energy was observed with a total energetic variation of 40-60 meV. This variation of optical bandgap is about one order of magnitude lower than the value used in our simulations. Calculations done with a fluctuation of ~3% of the nominal indium content (it roughly gives a variation of optical band gap of 5 meV), have not shown any noticeable difference from the case without fluctuations. It has been recently suggested that fluctuations along the in-plane direction of the quantum well could have greater magnitude and could be responsible for the enhancement of emission intensity due to the wavefunction localisation [20]. Our one-dimensional simulations lack the possibility of testing this idea although much work has been done towards extending the code to two dimensions.

6.2.9 Misfit dislocations

Another screening mechanism has been recently predicted. Rebane et al.[17] observed the presence of misfit dislocations within the InGaN quantum wells embedded in wurtzite GaN epilayers. They showed that misfit dislocations act as nonradiative centres and reduce the quantum efficiency of LEDs. On the other hand misfit dislocations partially release the strain inside the well and decrease the piezoelectric field. This leads to an increase in the optical matrix element for the electron-hole radiative recombination in the well and improves the quantum efficiency [17] of LEDs. Reference [17] showed that the optimal density of the misfit dislocations should be in the range 10^{5-6} cm^{-1}. In this case LED efficiency can improved by more than 10 times. Using our model described in chapter 2, these two effects of the misfit dislocations can easily be included. The value of piezoelectric polarization can be decreased to simulate an average strain-relaxation within the active region and the nonradiative recombination due to dislocations can be modelled by increasing the nonradiative recombination time in the SHR recombination equation 2.36 in chapter 2.

6.2.10 Excitons

It has been shown theoretically that the piezoelectric field causes a reduction of the exciton binding energy and of the absorption coefficient well below the values expected for a square quantum well. At higher carrier concentrations, many-body screening prevails over the screening of the built-in electric field, causing complete exciton bleaching [18]. This might explain why so far clear evidence of exciton bands is lacking in the photoluminescence excitation or absorption spectra of nitride quantum wells, in spite of the strong exciton binding energy of the bulk GaN and of the enhancement expected for the quantum size effect. For these reasons exciton transitions are not considered in the present work.

In conclusion, the fundamental mechanisms of spontaneous and stimulated light emission are still subject to debate. In particular, localisation effects due to composition fluctuations [19] and even quantum-dot-like structures due to phase

separation in InGaN quantum wells have been invoked to explain some of the observations [23][24]. More recently, it has been pointed out that many of the unusual optical properties of GaN-based heterostructure can be consistently explained by considering the piezoelectric properties of the III-nitrides [13][25][26]. We believe that piezoelectric field effects and spatial indium fluctuations must both be taken into account to understand the radiative recombination processes in InGaN quantum wells. Piezoelectric-field effects are likely to dominate in relatively thick wells (L >3 nm) of low indium concentrations (x <0.15-0.20). Indium composition fluctuation effects are likely to dominate in structures where the average indium concentration is high due to an increase in the magnitude of fluctuations [20]. Furthermore, misfit dislocations and traps due to the background doping can not be neglected for a comprehensive device simulation.

6.3 The CART device

Commonly, a single quantum well or multiple quantum wells are used as the active region in LEDs. For fabrication of a highly efficient device, the number of carriers recombining inside the active layer should be maximised, and the number of carriers recombining outside the active layer should be minimised. This needs optimisation of capture rates for electrons and holes into the active layer. In polar III-V and II-VI semiconductors the most effective channel of the carrier capture in quantum well is by emission of polar optical phonons. The corresponding carrier capture rate can be roughly estimated as the quantum well width divided by the product of the carrier thermal velocity and the polar phonon emission time. Thus, the capture rate depends on the quantum well parameters and the carrier masses. As a rule, in III-V semiconductors the electron effective masses are much lighter and corresponding thermal velocities are higher then those for holes. For this reason for a narrow quantum well, which provides optimal carrier confinement and maximises optical matrix elements, some of the electrons are not captured in the active layer and recombine outside it. This reduces the efficiency of LED devices.

To solve this problem, Wang et al.[2] suggested an LED structure based on the system of two wells with Charge Asymmetric Resonance Tunneling (CART), which allows enhancement of the number of the electrons captured into the ac-

tive layer. Here, a wide well (WW) and an active narrow quantum well (NW) are coupled via a narrow barrier which is transparent for electrons and blocking for holes. The high cross-section WW can more effectively capture electrons injected from the left. Then electrons tunnel into the NW and stay there until they combine radiatively with holes. Holes, coming from the right contact, have a sufficiently heavy mass that they can be captured even by a small cross-section NW. A blocking barrier towards the WW then confines holes in the NW. The phenomenon of the charge asymmetric resonance tunneling uses the quantum mechanical effect of strong exponential dependence of the potential barrier tunnel penetrability on the mass of the tunneling particle. By adjusting the alloy composition in the WW and NW and choosing an appropriate NW width, it is possible to match the energy position of the bottom of the WW with the energy position of the sub-band minimum of the NW, thus allowing for the possibility of electron resonant tunneling between the two wells. The heavy mass of the holes inhibits hole tunneling from the NW to the WW and the energy separation between the hole fundamental sub-band in the NW and WW means that hole penetration into the WW is not possible without thermal activation. Thus, the main effect of the coupled wells is to increase the quantum efficiency for the LED.

Growth of high quality InGaN quantum wells onto the GaN substrate is a difficult task in nitride technology [27][28], mainly because of the high lattice mismatch involved [19][29][13]. Consequently, high densities of non-radiative recombination defects and so low radiative efficiencies are commonly obtained at so-called dark spots. Moreover, the current associated with these dark spots causes serious overheating of the active region that reduces the device life time. In our structure, the NW is lattice matched onto the InGaN WW through the relaxed thin GaN barrier. In this way the WW works as a stopping layer for the threading dislocations due to the high stress at the n-GaN/InGaN interface. Therefore a good quality InGaN NW can be obtained. The WW acts also as a good current spreading layer, reducing the possibility of dark spots [2].

The CART LED design, shown schematically in figure 6.13(a), has produced bright blue LEDs [2].

Our results show that the piezoelectric field plays a key role into the correct functioning of the device and we outline the optimum conditions.

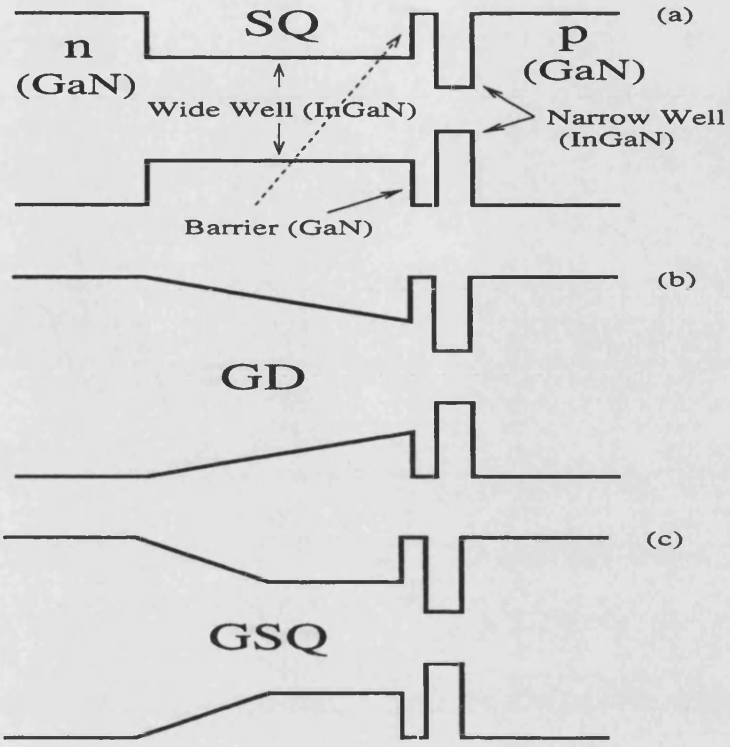

Figure 6.13: Schematic band profile for (a) the GaN-based CART LED with Square Wide Well, (b) the Graded Wide Well structure (GD), (c) the Graded and Square Wide Well structure (GSQ).

6.3.1 Simulation of the CART LED

One-dimensional drift-diffusion and Schrodinger equation solver has been used in our simulations. Initially a 2V bias was applied to a 1 μm device. The bulk p and n regions were biased to 10^{17} cm^{-3}. The 30 nm WW $In_{0.1}Ga_{0.9}N$, 2 nm GaN barrier and 3 nm $In_{0.2}Ga_{0.8}N$ NW were lightly n-doped at 10^{16} cm^{-3}. Our calculated band profile for the structure in figure 6.13(a) is shown in figure 6.14. First we have examined how to achieve resonant tunneling, namely the same bound state energy between the WW and NW at lower energy levels. Resonance between higher states is not sought as such states are less populated and thus the tunneling current is much reduced.

For a piezoelectric field of 0.55 MV/cm (right-hand panels of figure 6.14), E_{pz} and E_i cancel. As a result, the bottom of the WW is relatively flat. The ratio between the length of the WW and the length of the NW is critical for the ground states in the two wells to have equal values and has to be less than 10. This is because as the WW gets wider, its ground state approaches the bottom of the conduction

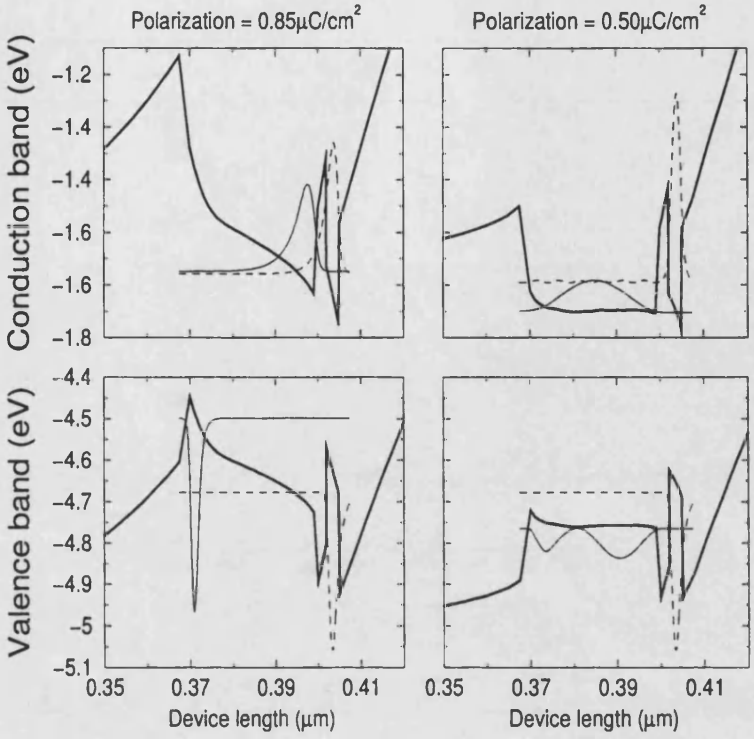

Figure 6.14: Band profiles of the device for two different values of the polarization charge density. Wavefunctions of the fundamental states are shown in the wide well (solid line) and in the narrow well (dashed line).

band and this is below the lowest energy state in the NW. This restriction can be overcome if x in the wide well is reduced.

The build up of charge due to a 0.94 MV/cm piezoelectric field at the junction between the n-type region and the WW is responsible for considerable bending of the right band profile of the device in the left-hand panels of figure 6.14. The band bending tends to raise the position of the ground state electron energy of the WW, keeping the corresponding state in the NW fixed. This can be used to achieve resonant tunneling despite the difference in widths of the WW and NW. At the same time, the ground state levels in the conduction and valence bands of the WW are moved apart, reducing recombination in the WW (which is proportional to the overlap of the wave functions). The first level in the conduction band of the WW for is moved closer to the barrier with the narrow well. In this way tunneling becomes more likely. In contrast, in the valence band the NW and WW ground states are spatially separated and so tunneling becomes negligible.

A barrier to the flux of electrons in the WW is a side effect of the polarization charge at the interface between the n-doped region and the WW. This barrier

reduces the electron current into the WW and so the efficiency of the device. To decrease the polarization, it is necessary to reduce the strain within the device, especially at the interface between n-doped bulk and WW. The idea is to grade x in the WW, thus reducing and spreading the strain across the well.

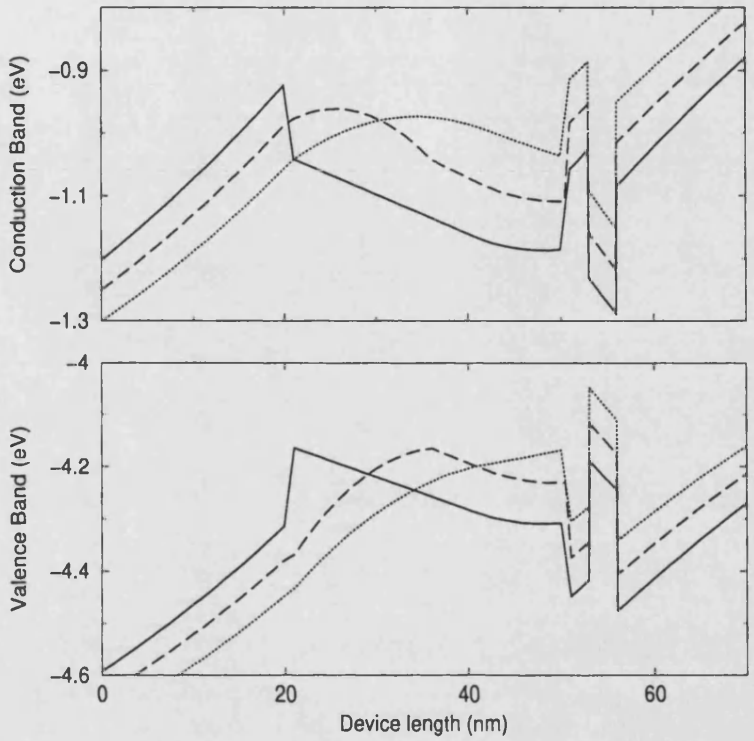

Figure 6.15: Band profile of square wide well (SQ)(solid), graded and square wide well (GSQ)(dotted) and graded wide well (GD)(dashed) structures.

Figure 6.15 shows the band profiles for the three different variants of figure 6.13: the solid line represents the square WW SQ design of figure 6.13(a). The dotted line shows the GSQ structure in which x in the WW is graded for the first half and then kept to a constant value as shown in figure 6.13(b). The dashed line describes the band profile for the fully graded WW (structure GD). Here x in the WW has a constant linear variation from 0 (at the junction with n-doped GaN) to 0.15 as illustrated in figure 6.13(c). The piezoelectric coefficients are taken from [8] scaled to 50% of their value. The choice of this value is critical and a long discussion was made in section 6.2. We found that this assumption gave the closest value in agreement with experiments reported in [2].

Since we vary x, a unique value of P_{pz} or E_{pz} cannot be given. Our simulation predicts a significant decrease in the electron barrier for structures GSQ and GD with the lowest value for structure GD. The latter has the lower strain as x of

0.15 is distributed across a 50nm region. The density of polarization charge is therefore lower and so is E_{pz}. While both structures GD and GSQ satisfy the requirement of a low electron injection barrier, the position of the wavefunctions is a deciding factor.

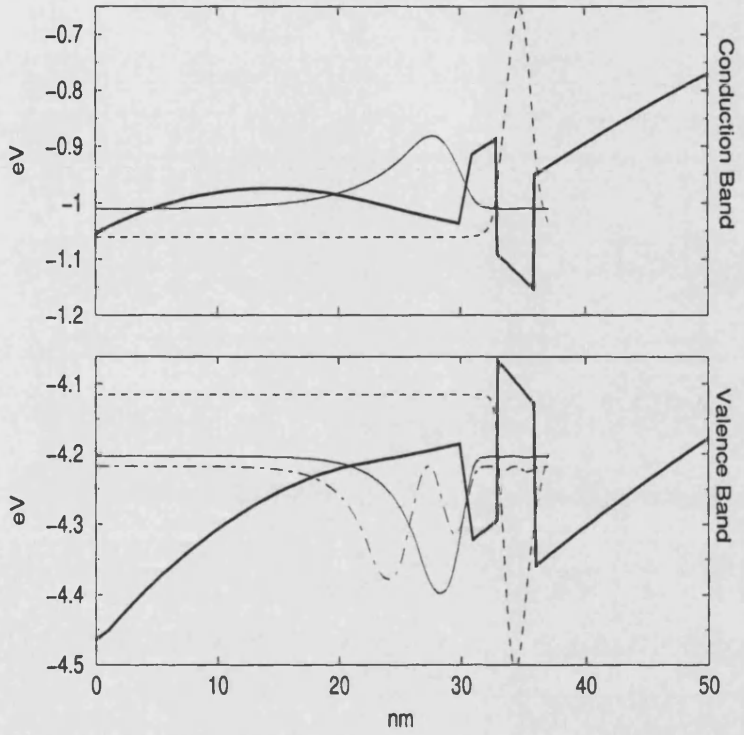

Figure 6.16: Wavefunctions of the GD structure: ground state in the NW (dashed line), ground state in the WW (solid line), first excited state in the WW (dot and dashed line).

Figure 6.16 shows the conduction and valence band profiles for the GD structure. The ground state energies are plotted for the WW and NW in the conduction band. In the valence band, a second energy level of the WW is shown. Even though the V_1 level in the NW of the valence band is far above the levels in the WW, the spatial separation between the wavefunctions is not as good as it was in the SQ structure. Rather, we can see that the wave function for the second level in the WW in the valence band has a noticeable tail inside the NW. Tunnelling is therefore possible. The GSQ structure, as shown in figure 6.17, conserves the tunnelling probability in the conduction band while reducing significantly the tunneling probability in the valence band.

In conclusion, our simulations show that the GSQ structure for the CART LED achieves a reduction in the electron injection barrier due to the piezoelectric field while maintaining optimal conditions for asymmetric tunneling.

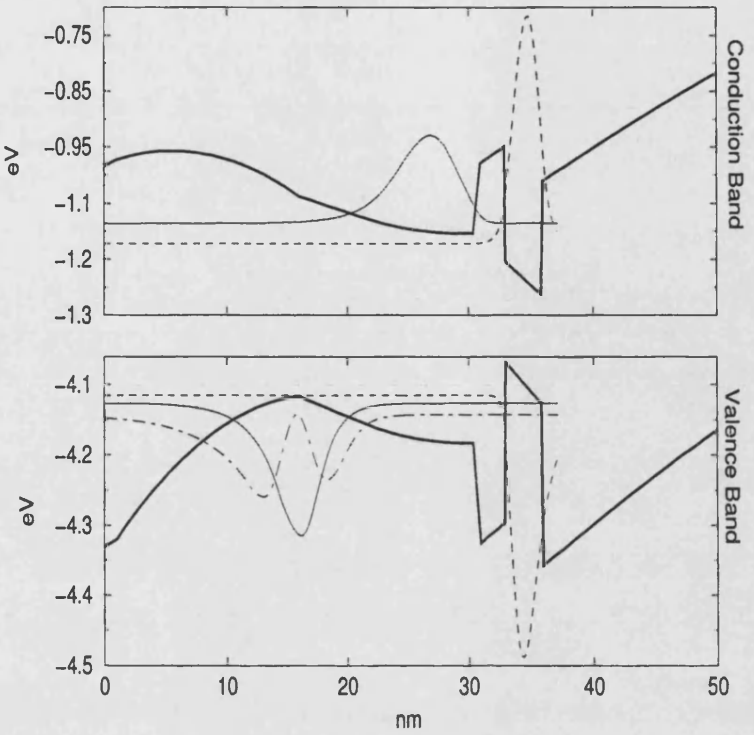

Figure 6.17: Wavefunctions in the GSQ structure: ground state in the NW (dashed line), ground state in the WW (solid line), first excited state in the WW (dot and dashed line).

6.3.2 Doping of the quantum wells

Laser structures based on III-V nitride multi-quantum wells (MQWs) have been shown to benefit from doping of the active region [30]. Experimentally, it has been established for GaN/InGaN heterostructures that doping of the active region produces a blue shift of the photoluminescence (PL) peak and an enhancement of the emission intensity [31]. It has been suggested that the observed phenomena are due to the enhancement by doping of the growth quality of the active region and screening of its potential fluctuations. Moreover, Di Carlo et al.[11] have shown how piezoelectric and spontaneous polarization fields can be screened via n- and p-type doping. Their simulations and other results [32][33][14][34] demonstrate that a multi-quantum well can be modulation doped to achieve screening of the macroscopic polarization field by the extrinsic carriers from ionised dopants.

A higher doping concentration in the bulk regions can help increase both the C_1-V_1 energy separation and the overlap. This is due to the fact that an effect of high doping concentrations is the reduction of the depletion region width and the consequent increase of the built-in electric field.

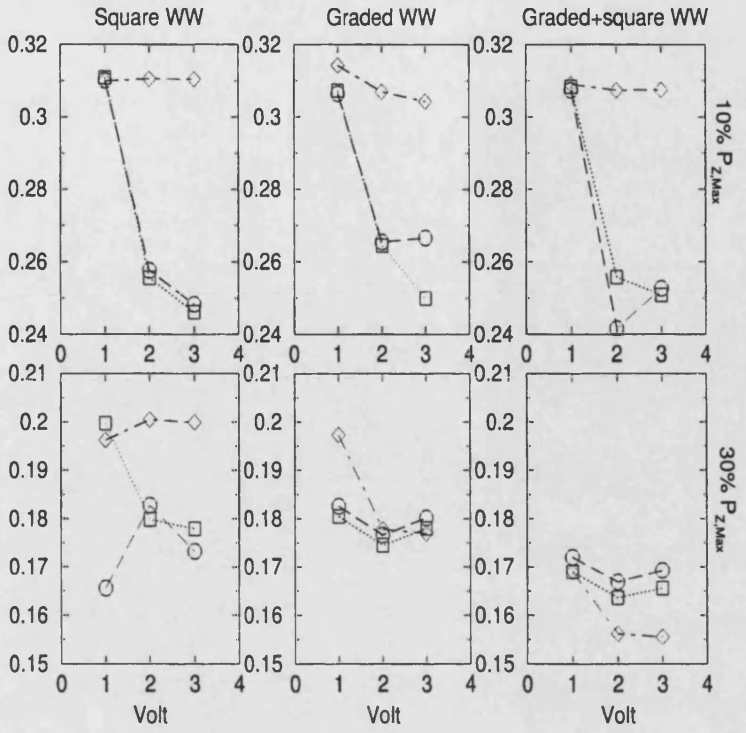

Figure 6.18: Overlap for the C_1-V_1 transition in the NW for three different structures: SQ (left panels), GD (centre panel) and GSQ (right panels). WW and NW have been doped with donor densities of 10^{15} cm^{-3} (circles), 10^{16} cm^{-3} (squares) and 10^{17} cm^{-3} (diamonds).

Figure 6.18 plots the overlap versus the device bias for three different doping densities and for the three different structures as before, for E_{pz}=0.3 MV/cm (top panels) and E_{pz}=1.3 MV/cm (bottom panels). As we can see, the high doping helps to increase the overlap when E_{pz} is small, whereas a large E_{pz} cancels most of the screening field by the dopant charges. The results can be explained from the fact that the fixed charge of ionised donors in both quantum wells creates an electric field. This field reduces the E_{pz}, and is responsible for the decrease in the QCSE and for the enhancement of the overlap between wave functions in the conduction and valence bands of the narrow quantum well as explained in section 6.2.3.

We have seen that for higher strains, a high doping concentration in the active region is required to increase the overlap. Unfortunately this leads to a large modification of the band profile as shown in Figure 6.19. The entire quantum structure formed by the wide and the narrow quantum well has been doped uniformly. Simulations have been carried out with two different values of doping. The higher value is of the same order of magnitude of the doping in the n- and p-regions. Figure 6.19 shows the band profile of the SQ structure for two biases.

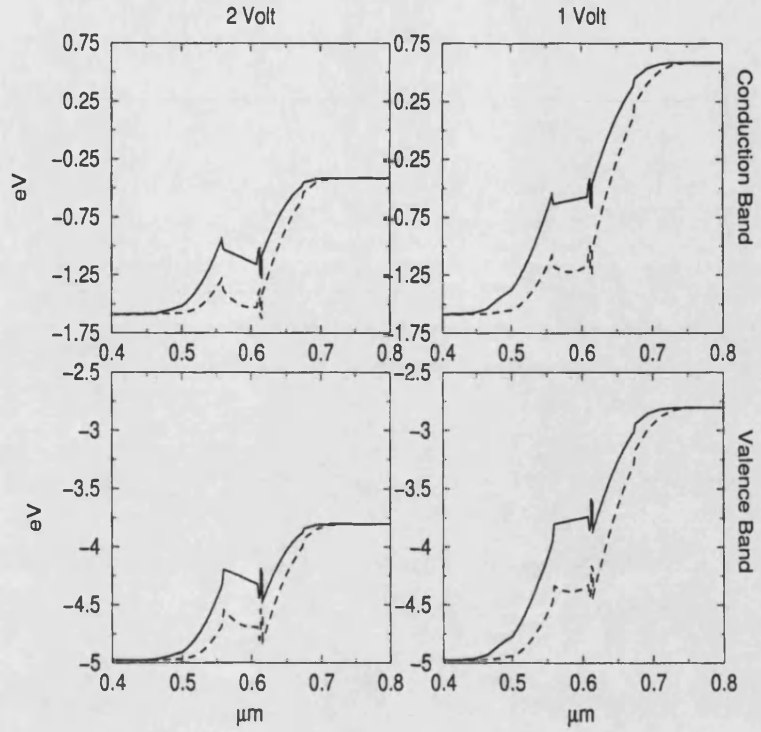

Figure 6.19: Band profile of the SQ structure for two biases: 2 V (left panels), 1 V (right panels). n-doping densities of 10^{15} cm^{-3} (solid line) and 10^{17} cm^{-3} (dashed line) have been used in the WW and NW.

The positive charge of the donors in the well helps screen E_{pz}; the bottom of the WW is flatter than for a low doping concentration. For this reason, p-doping is not a good solution, as the negative acceptor charge would create a field in the same direction as E_{pz}. With high doping concentrations, the WW and NW are shifted down towards the n-bulk region. The effect of this modification on electronic transport is large. Electrons are injected in the WW more easily while the hole current into the NW is much reduced and so is the efficiency of the device. This is due to the fact that the charge of ionised dopants is no longer negligible in comparison with the charge in the depletion region as the WW is typically wider than 50nm across.

6.4 Current-voltage characteristic

It is known [35] that the current-voltage characteristics of GaN-*pn* diode have distinct regions. The tunneling current dominates at low forward bias due to the presence of deep-level traps at the junction. A temperature-dependent diffusion

current is observed at moderate forward biases. In addition, space-charge-limited currents are found to be important at large forward and reverse biases. For LED functioning we are interested in the IV characteristic under a large forward bias. Simulation of a GaN *pn*-junction shows good agreement in the moderate bias region where drift-diffusion transport is dominant. For high biases, the space-charge-limited saturation of the current is seen only if traps are introduced. Figure 6.20 shows current-voltage characteristics for a 1μm GaN-*pn* junction (doping=10^{17} cm^{-3}) for three different densities and two depths of traps. We have used deep electron traps with energy levels of 0.5 eV and 0.25 eV below the conduction band energy level of GaN. The inset of figure 6.20 plots the IV curves with trap energy level of 0.5 eV for an extended range of bias. As we increase the trap density, the current saturates, a clear sign of the device entering the space-charge-limited current region. This behaviour is much more evident using deeper traps as shown by the comparison of dotted and solid lines in figure 6.20.

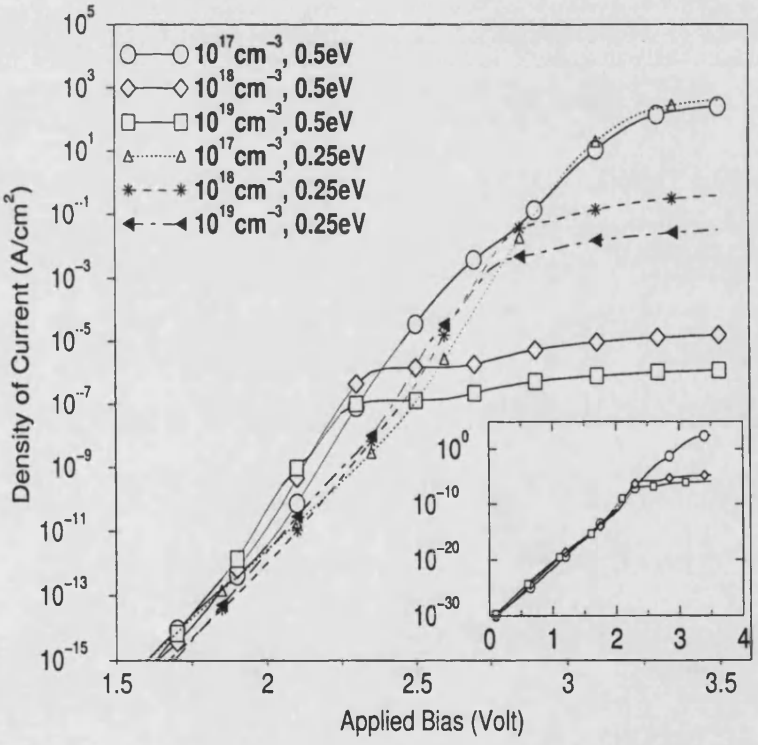

Figure 6.20: Current-Voltage curves for different trap densities and energy levels

I-V curves of a device with traps at an energy level of 0.5 eV below the conduction band have a saturation point just above a bias of 2 V. Reducing the distance of the trap energy level from the conduction band, the saturation point shifts to higher biases in good agreement with experiments[35],[36]. There is still uncertainty concerning the nature of these traps . We believe that the background *n*-type

doping that has a density of $5{\cdot}10^{17}$ cm^{-3} could be a valid candidate. Also misfit dislocations(density $\sim 10^{9-10}$ cm^{-2}) could be used to explain the trap behaviour of the I-V characteristic.

6.5 Conclusions

We have investigated in depth the effect of the piezoelectric field on the performance of the device. Matching the experimental PL data we are able to predict the size of the piezoelectric field and the resulting band profile of the device. We have shown that there is strong uncertainty on the values of piezoelectric constants found in the literature. We argue that conditions of growth of the device are crucial in affecting the value of the piezoelectric field to the extent that a fixed value of this field is not applicable to every device. We have described a new GaN/InGaN LED that, by using resonant tunneling between two quantum wells, potentially achieves a higher quantum efficiency than conventional single quantum well devices. We have simulated this device using our drift-diffusion model coupled with self-consistent solution of Poisson's and Schrödinger's equation. Finally, we have shown that to explain the experimental I-V characteristic we need to include an high density of traps in an GaN-*pn*-junction. This density is consistent with the existence of high background *n*-type doping and misfit dislocations.

References

[1] J. Fischer, H. Kuhne and H. Richter, *Phys. Rev. Lett.*, **73**, 2712 (1994).

[2] Y. T. Rebane, Y. G. Shreter, B. S. Yavich, V. E. Bougrov, S. I. Stepanov and W. N. Wang, *Phys. Stat. Sol.*, **180**, 121 (2000).

[3] C. Wetzel, S. Nitta, T. Takeuchi, S. Yamaguchi, H. Amano and I. Akasaki, *MRS Internet J Nitride Semicond. Res.*, *3*, 31 (1998).

[4] I. Akasaki and H. Amano, *Jpn. J. Appl. Phys.*, **36**, 5393 (1997).

[5] S. I. Stepanov, W. N. Wang, B. S. Yavich, V. E. Bougrov, Y. T. Rebane and Y. G. Shreter, *MRS Internet J. Nitride Semicond. Res.*, **6**, 6 (2001).

[6] T. Takeuchi, C. Wetzel, S. Yamaguchi, H. Sakai, H. Takeuchi, H. Amano, I. Akasaki, Y. Kaneko, S. Nakagawa, Y. Yamaoka and N. Yamada, *Appl. Phys. Lett.*, **73**, 1691 (1998).

[7] S. F. Chichibu, T. Sota, K. Wada, S. P. Denbaars and S. Nakamura, *MRS Internet J. Nitride Semicond. Res.*, **4S1**, G27 (1999).

[8] F Bernardini, V Fiorentini, D Vanderbilt, Phys Rev B, **56**, R10024, (1997).

[9] C. Wetzel, T. Takeuchi, H. Amano and I. Asaki, *Phys. Rev. B*, **62**, R13302 (2000).

[10] P. Perlin, C. Kisielowski, V. Iota, B. A. Weinstein, L. Mattos, N. A. Shapiro, J. Kruger, E. R. Weber and J. Yang, *Appl. Phys. Lett.*, **73**, 2778 (1998).

[11] A. Di Carlo, F. Della Sala, P. Lugli, V. Fiorentini and F. Bernardini, *J. Appl. Phys.*, **87**, 2285 (2000).

[12] C. Wetzel, T. Takeuchi, H. Amano and I. Akasaki, *MRS Internet J. Nitride Semicond. Res.*, **4S1**, G366 (1999).

[13] A. Hangleiter, J. S. Im, H. Kollmer, S. Heppel, J. Off and F. Scholz, *MRS Internet J Nitride Semicond. Res.*, **3**, 15 (1998).

[14] A. N. Cartwright, P. M. Sweeney, T. Prunty, D. P. Bour and M. Kneissl, *MRS Internet J. Nitride Semicond. Res.*, **4**, 12 (1999).

[15] A. D. Bykhovski, V. V. Kaminski, M. S. Shur, Q. C. Chen and M. A. Khan, *Appl. Phys. Lett.*, **68**, 818 (1996).

[16] N. A. Shapiro, P. Perlin, C. Kisielowski, L. S. Mattos, J. W. Yang and E. R. Weber, *MRS Internet J. Nitride Semicond. Res.*, **5**, 1 (2000).

[17] Y. T. Rebane, Y. G. Shreter and W. N. Wang, *Appl. Surf. Science*, 6313 (2000).

[18] G. Traetta, R. Cingolani, A. Di Carlo, F. Della Sala, P. Lugli, *Appl. Phys. Lett.*, **76**, 1042 (2000).

[19] M. B. Nardelli, K. Rapcewicz and J. Bernholc, *Appl. Phys. Lett.*, **71**, 3135 (1997).

[20] N. A. Shapiro, Y. Kim, H. Feick, E. R. Weber, P. Perlin, J. W. Yang, I. Akasaki and H. Amano, *Phys. Rev. B*, **62**, R16318 (2000).

[21] A. D. Bykhovski, B. Gelmont and M. Shur, *J. Appl. Phys.*, **74**, 6734 (1993).

[22] A. Bonfiglio, M. Lomascolo, G. Traetta, R. Cingolani, A. Di Carlo, F. Della Sala, P. Lugli, A. Botchkarev and H. Morkoc, *J. Appl. Phys.*, **87**, 2289 (2000).

[23] S. Chichibu, T. Azuhata, T. Sota, S. Nakamura, *Appl. Phys. Lett.*, **69**, 4188 (1996).

[24] Y. Narukawa, Y. Kawakami, S. Fujita, S. Nakamura, *Phys. Rev. B*, **55**, R1938 (1997).

[25] T. Takeuchi, S. Sota, M. Katsuragawa, M. Komori, H. Takeuchi, H. Amano, I. Akasaki, *Jpn. J. Appl. Phys.*, **36**, L382 (1997).

[26] J. S. Im, H. Kollmer, J. Off, A. Sohmer, F. Scholz, A. Hangleiter, *Phys. Rev. B*, **57**, R9435 (1998).

[27] S. C. Jain, *Appl. Phys. Rev.*, **87**, 965 (2000).

[28] S. J. Pearton, J. C. Zolper, R. J. Shul and F. Ren, *Appl. Phys. Rev.*, **86**, 1 (1999).

[29] M. B. Nardelli, K. Rapcewicz and J. Bernholc, *Phys. Rev. B*, **55**, R7323 (1997).

[30] S. Nakamura, *Appl. Phys. Lett.*, **70**, 2753 (1997).

[31] H. Jiang, M. Minsky, S. Keller, E. Hu, J. Singh and S. P. DenBaars, *IEEE J. Quantum Electronics*, **35**, 1483 (1999).

[32] P. Kozodoy, Y. P. Smorchkova, M. Hansen, H. Xing, S. P. DenBaars, U. K. Mishra A. W. Saxler, R. Perrin and W. C. Mitchell, *Appl. Phys. Rev.*, **75**, 2444 (1999).

[33] F. Della Sala, A. Di Carlo, P. Lugli, F. Bernardini, V. Fiorentini, R. Scholz and J. Jancu , *Appl. Phys. Lett.*, **74**, 2002 (1999).

[34] V. Fiorentini and F. Bernardini, F. Della Sala, A. Di Carlo, P. Lugli, *Phys. Rev. B*, **60**, 8849 (1999).

[35] J. B. Fedison, T. P. Chow, H. Lu and I. B. Bhat, *Appl. Phys. Lett.*, **72**, 2841 (1998).

[36] F. Calle, E. Monroy, F. J. Sanchez, E. Munoz, B. Beaumont, S. Haffouz, M. Leroux and P. Gibart, *MRS Internet J Nitride Semicond. Res.*, **3**, 24 (1998).

Chapter 7

Conclusion

The preceding examples demonstrate the powerful engineering utility of simple hydrodynamic transport equations coupled with a self-consistent solution of Poisson's and Schrödinger's equations, which serve to govern a physical model of TEDs and LEDs. This model is based on several phenomenological parameters derived from Monte Carlo particle simulations, and, therefore, the fundamental aspects of the semiconductor transport physics are built into simpler, tractable questions. Since the computer program is quite general, it should be noted that transport parameters from semiconductor materials other than GaAs and GaN can be used as well. Moreover, characteristics typical of III-nitrides such as strain and piezoelectric polarization are built in the model. Several orders of magnitude reduction in computer costs and resources over the Monte Carlo method can be realised permitting efficient and systematic study of device characteristics.

For the simulation of Gunn diodes, the program has been tested with Monte Carlo results and we have obtained good agreement with the literature, in particular an almost perfect matching of results with previous works utilising a similar Drift-Diffusion approach. We have concentrated our analysis mainly in the study of two particular devices. A simple n-layer structure has shown the importance of the boundary conditions for the mode of propagation of a Gunn diode. This has been useful in understanding of the dynamic of the second device: a graded-gap AlGaAs heterostructure Gunn diode. It has been explained that it is possible to improve the performance of a Gunn diode by just modifying the cathode region. Our simulations show the effects of insertion of a n^+ spike in graded-gap AlGaAs

Gunn diode: the spike creates a charge sheet necessary for tailoring the electric field in the transit region and separating the injector from the transit region. In addition, with our drift-diffusion model, we have been able for the first time to understand how the spike works and the limits of its use. We have proved that compensation of charge within the device is a key issue.

We have investigated in depth the effect of the piezoelectric field on the performance of GaN-based LEDs. Matching the experimental PL data we are able to predict the size of the piezoelectric field and the resulting band profile of the device. We have shown that there is strong uncertainty on the values of piezoelectric constants found in the literature. We argue that conditions of growth of the device are crucial in affecting the value of the piezoelectric field to the extent that a fixed value of this field is not applicable to every device. We have described a new GaN/InGaN LED that, by using resonant tunneling between two quantum wells, potentially achieves a higher quantum efficiency than conventional single quantum well devices. We have seen that the device performances are deeply affected by the piezoelectric field and alternative designs have been proposed. A graded-square well design has been the more promising solution. We have investigated the effect of doping in the active region and found out that screening of piezoelectric field is achieved at the expenses of a reduction in the hole injection efficiency.

Finally, we have shown that to explain the experimental I-V characteristic of GaN-*pn*-junctions, we need to include a high density of traps. This density is consistent with the existence of high background n-type doping and misfit dislocations.

7.1 Future work

The idea that has motivated our work is that a simple drift-diffusion model can still be used in the simulation of modern semiconductor devices. However in order to obtain a complete description of physical phenomena we needed to couple this simple model with others tools such as Monte Carlo-derived parameters or Schrödinger equation solvers. This improvement of the drift diffusion method with other techniques is what we believe is the major future task.

In the simulation of heterostructures Gunn diode, we believe that the natural next step could be the inclusion of energy transport equations in the model. This improvement would help to obtain a more rigorous description of hot electron effects in the device.

As far as GaN LEDs are concerned, there are many problems that have to be tackled in the future. First of all, there is the issue of the second dimension. Many effects such as alloy fluctuations, misfit or threading dislocations in the quantum well have a two-dimensional nature. There are not major difficulties since it is just a matter of increase the number of equations solved. However the increase in computational effort needs to be taken in consideration. At the present, the Schrödinger solver is the part of the computer code that affects the computational speed most heavily. This solver is greatly adaptive and can be used for almost any kind of potential profile. Unfortunately it has intrinsic stability problems due to the transfer matrix technique used. Different methods could be used, although an improvement of the present technique is, in our opinion, preferable. A second important issue is the degree of approximation that has been made regarding the band structure of nitrides. Under strain, the valence band changes so dramaticaly that the parabolic-band assumption is no longer acceptable. The effective mass approximation must be reconsidered especially in proximity to the quantum well. The problem of coupling the classic transport with quantum regions such as a quantum well must be analysed more deeply. The inclusion of a proper optical recombination mechanism in the quantum well and in the bulk will allow us to predict the emission spectrum, making easier the comparison of simulation and experimental results. The work on GaN is only at its early stage; much remains to be done. After LEDs, laser diodes could be simulated using rate equations and non-equilibrium dynamical models.

Validation of the models must be the main priority. In this case, experiments need to be done in order to verify the accuracy of simulations.

Aside from this, we wish to emphasise the necessity of improving the speed and efficiency of the computer algorithms, particularly if the extension to a 2-dimensional simulation is desired.

Semiconductor device modelling remains one of the most important tools available to researchers who wish to gain knowledge of the physics of semiconductor

devices. However, device modelling can not be done without a high quality theory of semiconductor devices. Furthermore, validation of the simulations must be achieved through experimental tests.

We believe that the real purpose of simulation is that you get enough insight so that you never have to run the simulation again.

Chapter 8

APPENDICES

8.1 Appendix A

GaAs-AlGaAs Semiconductors

Gallium Arsenide is a direct gap semiconductor. Its band structure is shown in figure 8.1. The valence band consists of four subbands when spin is neglected and each band is doubled when spin is taken into account. Three of the four bands are degenerate at k=0 (Γ-point) and form the upper edge of the band. The fourth band forms the bottom. Furthermore, the spin-orbit interaction causes a splitting of the band at k=0. As shown in figure 8.1 along a given direction the two top valence bands are approximately near the k=0 and can be fitted by two parabolic bands with different curvatures: the heavy-hole band and the light-hole band. The conduction band consists of a number of subbands and has its minimum at k=0. For high-purity gallium arsenide, at room temperature and under normal atmosphere, the value of the direct band gap is around 1.424 eV. Others parameters are reported in table 8.1. The bowing parameter for $Al_xGa_{1-x}As$ is b=0.62 [3] defined by:

$$E_g(x) = E_{AlAs}x + E_{GaAs} - b(1-x)x \tag{8.1}$$

The difference of energy gap between AlAs and GaAs is split between conduction and valence band as 65:35 [4]. Same ratio is assumed valid for $Al_xGa_{1-x}As$/GaAs

interface.

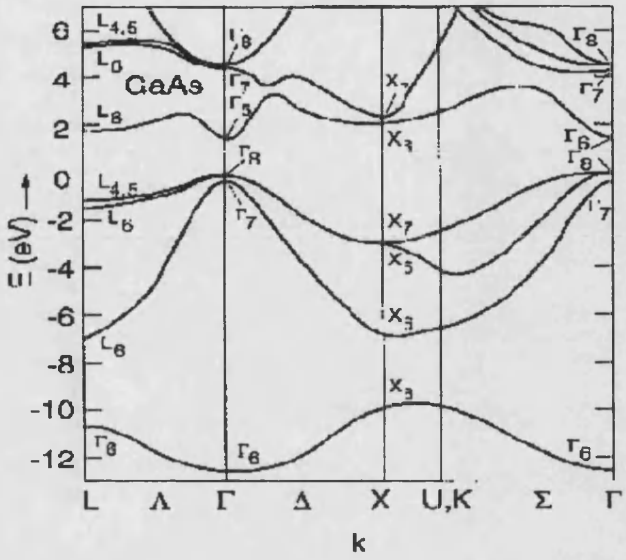

Figure 8.1: Band Structure of GaAs [1]

References

[1] J. S. Blakemore, *Key Papers in Physics-Gallium Arsenide*, American Institute of Physics, New York (1987).

[2] S. Adachi, *Physical Properties of III-V Semiconductor Compounds*, John Wiley and Sons, New York, 1992.

[3] S. R. Lee, A. F. Wright, M. H. Crawford, G. A. Petersen, J. Han and R. M. Biefeld, *Appl. Phys. Lett.*, **74**, 3344 (1999).

[4] H. Kroemer,*Surf. Sci.*, **174**, 299 (1985).

Name	Symbol	Value	Unit
Effective density of states in conduction band	N_C	$1.675{\cdot}10^{18}$	cm^{-3}
Effective density of states in valence band	N_V	$1.2{\cdot}10^{19}$	cm^{-3}
Electronic affinity	χ_c	4.7	eV
Energy gap	E_g	1.424	eV
Relative dielectric constant	ϵ_r	13.18	
Electron mobility	μ_n	8500	cm^2/Vs
Hole mobility	μ_p	300	cm^2/Vs
Fraction of electron effective mass	m_e	0.067	
Fraction of hole effective mass	m_h	0.5	
Electron recombination time	τ_n	$1{\cdot}10^{-9}$	s
Hole recombination time	τ_n	$1{\cdot}10^{-9}$	s
Ionisation donor energy	E_D	$1{\cdot}10^{-2}$	eV
Ionisation acceptor energy	E_A	$1{\cdot}10^{-2}$	eV

Table 8.1: GaAs material parameters [2]

Name	Symbol	Value	Unit
Effective density of states in the conduction band	N_C	$\cdot 10^{18}$	cm^{-3}
Effective density of states in the valence band	N_V	$\cdot 10^{19}$	cm^{-3}
Electronic affinity	χ_c		eV
Energy gap	E_g	2.15	eV
Relative dielectric constant	ϵ_r	10.1	
Electron mobility	μ_n	8500	cm^2/Vs
Hole mobility	μ_p	300	cm^2/Vs
Fraction of electron effective mass	m_e	0.15	
Fraction of hole effective mass	m_h	0.5	
Electron recombination time	τ_n	$1{\cdot}10^{-9}$	s
Hole recombination time	τ_n	$1{\cdot}10^{-9}$	s
Ionisation donor energy	E_D	$1{\cdot}10^{-2}$	eV
Ionisation acceptor energy	E_A	$1{\cdot}10^{-2}$	eV

Table 8.2: AlAs material parameters [2]

8.2 Appendix B

Crystal structure

The common semiconductors form highly symmetric crystals, and their symmetry is an essential ingredient in advanced theory. For example , it is often possible to show which optical processes are allowed or forbidden from the symmetry alone, without detailed knowledge of the band structure or wave functions. Group theory provides the mathematical framework for analyzing symmetry [1]. In this section we will give few notions on crystals and we will put down the basis for understanding the theory of piezoelectric materials. For further insight on the crystal theory refer to [2].

The most common semiconductors like Si and GaAs have cubic crystals. Cubic lattices are shown in figure 8.2(a)-(d). The simple cubic lattice is primitive, meaning that it has only one atom per unit cell of volume a^3. A further atom can be added in the middle of each cube, giving a body-centred cubic (BCC) lattice with two atoms, or at the centre of each face to give a face-centred cubic (FCC) lattice with four atoms in each unit cell.

The common semiconductors are based on the FCC structure but with twice the number of atoms. The structure of elemental semiconductors such as Si is known as the *diamond* lattice is shown in Figure 8.2(d). There is an additional atom displaced by $(\frac{1}{4}, \frac{1}{4}, \frac{1}{4})a$ with respect to each atom in the FCC lattice. This correspond to the chemical view that each atom forms four sp^3 bonds to its neighbours. The cubic unit cell of diamond contains eight atoms. Compounds such as GaAs have the *zinc-blende*. The arrangement of the atoms is the same as in the diamond but the two species alternate: Ga occupies the original sites of the FCC lattice, while As occupies the tetrahedral sites, to give a unit cell that contains four atoms of each species.

There is a standard notation for describing planes and directions in crystals. Directions are specified by writing the cartesian components of a vector without commas inside square brackets. For example the z-axis is [001]. Planes are denoted in a similar way by *Miller indices*, writing a normal to the plane in parentheses, such as (001) for the xy-plane. The notation 001 means the plane (001) and all those related by symmetry, such as (100) in a cubic lattice. The

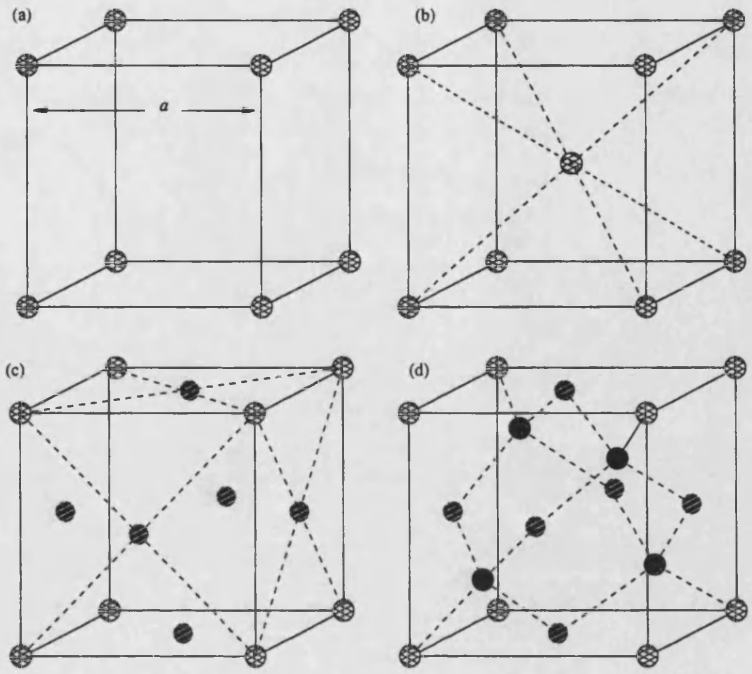

Figure 8.2: Cubic crystal structures: (a)simple, (b)body-centred (BCC), and (c) face-centred cubic lattices(FCC). The structure of the common semiconductors is (d) diamond for elements or zinc-blende for compounds.

notation is intended to describe sets of equally spaced planes. If one such plane passes through the origin, the set of planes is described by the reciprocals of the intercepts of the next plane on the axes, in units of the lattice constant [3].

The lower symmetry of compounds has some important consequences. For example, they are piezoelectric, which means that certain strains generate an electric field. Compressing the lattice along the [111]-direction will change the relative spacing of the planes of Ga and As. The two species of atom are not precisely neutral but carry opposite electric charge. Thus a relative displacement of the planes sets up a polarization $\mathbf{P}$, which may in turn lead to an electric field. Not all strains have this effect; compression along [001], for example, causes no polarization.

In the zinc-blende structure, a view along a [111]-direction reveals that the atoms of a given kind are stacked in the sequence $...ABCABC...$ while maintaining tetrahedral bonds with those of the other kind. The underlying Bravis lattice is cube. There is another arrangement that preserves the tetrahedral bonding but stacks the atoms of a given kind in the sequence $...ABABAB...$ this is the

wurtzite structure, whose underlying Bravais lattice is hexagonal. The wurtzite structure, typical for example of nitride compounds such as GaN, is shown in figure 8.3. This lattice consists of hexagonal closest packed gallium ions with the nitride ions occupying 1/2 of the tetrahedral holes.

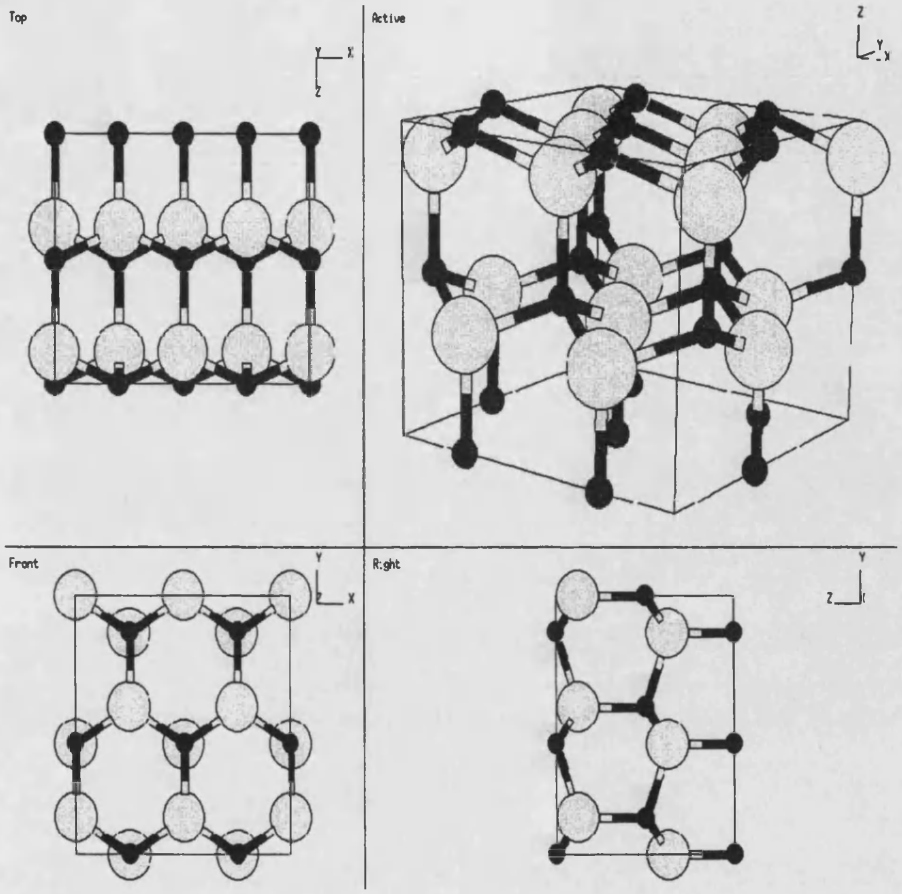

Figure 8.3: Wurtzite structure from several perspectives [1].

8.2.1 Polarization

The relative permittivity of a material is a directly measurable quantity which expresses on a macroscopic scale the overall result of the interaction which occurs on a macroscopic scale between an externally applied field and the atoms or molecules of the material. This interaction is referred to as polarization.

Insulators or dielectrics are material whose electrons are bound to the atomic structure by such strong forces that they are not free to move throughout the

material under the influence of an applied electric field. The electrons can, however, be displaced very small distances relative to their nuclei.

Dielectrics may be broadly divided into non-polar materials and polar materials. In non-polar materials the molecules, which are usually diatomic and composed of two atoms of the same type, may be represented as positive nuclei of charge q surrounded by a symmetrically distributed negative electron cloud of charge $-q$. In the absence of an applied field the centres of gravity of the positive and negative charge distributions coincide. When the molecules are placed in an external electric field the positive and negative charges experience electric forces tending to move them apart in the direction of the external field. The centres of positive and negative charges no longer coincide and the molecules are said to be polarized. Each molecule now forms a dipole whose moment is defined as

$$\mathbf{p} = q\mathbf{dl} \tag{8.2}$$

where $\mathbf{dl}$ is the distance between the two centres of charge and is a vector pointing from the negative to the positive charge. Dipoles so formed are known as induced dipoles since when the field is removed the charges resume their normal distribution and the dipoles disappear.

References

[1] A. W. Joshi, *Elements of Group Theory for Physicists*, Wiley Eastern Limited, Third Edition, 1982.

[2] N. W. Ashcroft and N. D. Mermin, *Solid State Physics*, Saunders (1976).

[3] J. H. Davies, *The Physics of Low-Dimensional Semiconductors*, Cambridge University Press, 1998.

8.3 Appendix C

Solutions of algebraic linear systems

The numerical solution of large systems of algebraic equations is a direct consequence of the Finite Difference method for solving ODEs or PDEs. The goal in these techniques is to break the continuous differential equation into a coupled set of algebraic difference equations for each finite volume or node in the system. When one has only a single independent variable (the ODE case), this process can easily lead to several hundred simultaneous equations that need to be solved. For multiple independent variables (the PDE case), systems with hundreds of thousands of equations are common. Thus, in general, we need to be able to solve large systems of linear equations of the form $\mathbf{Ax} = \mathbf{b}$ as part of the solution algorithm for general Finite Difference methods.

There are two general schemes for solving linear systems: Direct Elimination Methods and Iterative Methods. All the direct methods are based on the standard Gauss Seidel Elimination technique, which systematically applies row operations to transform the original system of equations into a form that is easier to solve. This appendix summarizes a few algorithms for implementation of the basic Gauss Elimination scheme and it also highlights the LU Decomposition method which, although functionally equivalent to the Gauss Elimination method, does provide some additional flexibility for computer implementation. Thus, the LU decomposition method is often the preferred direct solution method for low to medium sized systems (usually less than 200-300 equations).

For large systems, iterative methods (instead of direct elimination methods) are almost always used. This switch is required from accuracy considerations (related to round-off errors), from memory limitations for physical storage of the equation constants, from considerations for treating nonlinear problems, and from overall efficiency concerns. There are several specific iterative schemes that are in common use, but most methods build upon the base Gauss Seidel method, usually with some acceleration scheme to help convergence.

8.3.1 Direct Methods

The Gauss Elimination Method forms the basis for all elimination techniques. The basic idea is to modify the original equations $\mathbf{Ax} = \mathbf{b}$, using row operations, to give a simpler form for actual solution. The basic algorithm can be broken into two stages: forward elimination (put equations in upper triangular form) and back substitution (solve for unknown solution vector), details of this method can be found on every text book [1].

The Gauss Elimination Method has the disadvantage that all right-hand sides (i.e. all the vectors of interest for a given problem) must be known in advance for the elimination step to proceed. The LU Decomposition Method outlined here has the property that the matrix modification (or decomposition) step can be performed independent of the right hand side vector. This feature is quite useful in practice, therefore, the LU Decomposition Method is usually the direct scheme of choice in most applications.

To develop the basic method, we break the coefficient matrix into a product of two matrices,

$$\mathbf{A} = \mathbf{LU} \tag{8.3}$$

where $\mathbf{L}$ is a lower triangular matrix and $\mathbf{U}$ is an upper triangular matrix. Now, the original system of equations, $\mathbf{Ax} = \mathbf{b}$ can be divided into two problems,

$$\mathbf{Ly} = \mathbf{b} \qquad \mathbf{Ux} = \mathbf{y} \tag{8.4}$$

The rationale behind this approach is that the two systems given in equation 8.4 are both easy to solve; because $\mathbf{L}$ is a lower diagonal matrix, the expression,$\mathbf{Ly} = \mathbf{b}$, can be solved with a simple forward substitution step. Similarly, since $\mathbf{U}$ has upper triangular form, $\mathbf{Ux} = \mathbf{y}$ can be evaluated with a simple back substitution algorithm.

Thus the key to this method is the ability to find two matrices, $\mathbf{L}$ and $\mathbf{U}$, that

satisfy equation 8.3. Doing this is referred to as the Decomposition Step and there are a variety of algorithms available. Three specific approaches are as follows: Doolittle Decomposition, Crout Decomposition and Cholesky Factorization. For a detailed description of this methods refer to [1].

8.3.2 Iterative Methods

For large systems of equations, an iterative solution scheme for the unknown vector can always be written in the form

$$\mathbf{x}^{p+1} = \mathbf{B}\mathbf{x}^{p} + \mathbf{c} \tag{8.5}$$

where $\mathbf{B}$ is the iteration matrix, $\mathbf{c}$ is a constant vector and p is an iteration counter. Convergence of this scheme is guaranteed if the largest eigenvalue of the iteration matrix is less than unity, where

$$\rho = spectral \quad radius = |\lambda_{Max}| \tag{8.6}$$

Therefore, if $\rho < 1$ the iterative scheme will converge. If $\rho << 1$, the iterative scheme converges very rapidly. If $\rho \sim 1$ but less than unity, the scheme will be slowly converging. The iteration algorithm will diverge if the spectral radius is greater than unity.

Convergence is tested during the iterative process by computing the largest relative change from one iteration to the next, and comparing the absolute value of this result with some desired tolerance. If the maximum relative change is less than the desired accuracy, then the process is terminated. If this condition is not satisfied, then another iteration is performed.

The Gauss Seidel Method

The original system of equations given by $\mathbf{Ax} = \mathbf{b}$ can be convert it into the classical Gauss Seidel iterative scheme [1]:

$$x_i^{p+1} = \frac{1}{a_{ij}} \left(b_i - \sum_{j=1}^{i-1} aijx_j^{p+1} - \sum_{j=i+1}^{N} aijx_j^p \right). \tag{8.7}$$

Thus, if the equations are taken in sequence, all the terms on the right hand side are known, thereby allowing computation of a new estimate for the full solution vector in terms of the very latest information available for the calculation.

The Successive Over-Relaxation (SOR) Method

To improve the rate of convergence, one might consider using a weighted average of the results of the two most recent estimates to obtain the next best guess of the solution. If the solution is converging, this might help extrapolate to the real solution more quickly. This idea is the basis of the SOR method.

In particular, let α be some weight factor with a value between 0 and 2. Now, we compute the next value of $\mathbf{x}^{p+1}$ to use in the Gauss Seidel method as a linear combination of the current value, $\mathbf{x}^{p+1}$, and the previous solution,$\mathbf{x}^p$, as follows:

$$\mathbf{x}^{p+1}|_{new} = \alpha\mathbf{x}^{p+1} + (1-\alpha)\mathbf{x}^p \qquad 0 < \alpha < 2 \tag{8.8}$$

Note that if α is unity, we simply get the standard Gauss Seidel method (or whatever base iterative scheme is in use). When α is greater that unity, the system is said to be over-relaxed, indicating that the latest value,$\mathbf{x}^{p+1}$, is being weighted more heavily (weight for $\mathbf{x}^p$ is negative). If, however, α is less than one, the system is under-relaxed, this time indicating that the previous solution,$\mathbf{x}^p$, is more heavily weighted (positive weight values).

The idea is to choose the relaxation parameter to improve convergence (reduce

the spectral radius). This is most often done in a trial and error fashion for certain classes of problems (experience helps here). Our code tries to estimate this quantity as part of the iterative calculation.

References

[1] W.H Press et al, *Numeical Recipies*, Cambridge University Press.

8.4 Appendix D

The Transfer Matrix Technique

This appendix contains a brief explanation of the Transfer Matrix Technique (TMT). For a deeper insight about this technique, refer to J.H Davies *The Physics of Low-Dimensional Semiconductors* edited by Cambridge University Press (1998).

We consider an electron of energy E, travelling in the positive x-direction. We suppose that at $x = 0$ the electron is scattered by a potential step as shown in figure 8.4,

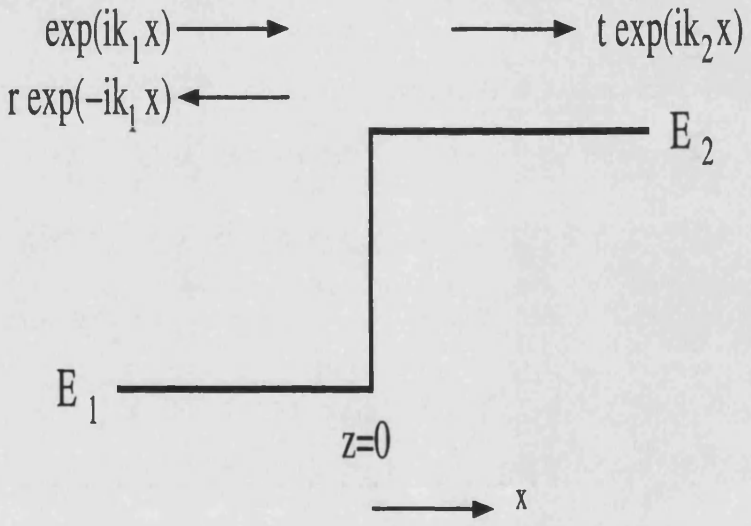

Figure 8.4: Potential energy profile seen by an electron travelling in the positive x-direction. A reflected wave on the left and a transmitted wave on the right is formed after the scattering process.

E_1 and E_2 are the band energies before and after the step, respectively. Classically the electron would always be reflected if its energy $E < E_2$, whereas it always passes over the step if $E > E_2$. Quantum mechanics, instead, predict the formation of a transmitted and reflected electron for all the values of $E_1 < E < E_2$.

Incoming waves $\exp(ik_1x)$ in region 1 generates two outgoing waves, a reflected wave $r\ \exp(-ik_1x)$ in region 1 and a transmitted wave $t\ \exp(ik_2x)$ in region

2. The temporal factor of $\exp(-iEt/\hbar)$ is the same for all waves and will be ignored. The coefficient r and t are reflection and transmission amplitudes and are complex in general. The wave numbers are given by $k_1^2 = 2m(E - E_1)/\hbar^2$ and $k_2^2 = 2m(E - E_2)/\hbar^2$. If we assume first that $E > E_2$, k_2 is real and all the waves propagate. We generalise to the situation where there is an incoming wave from the right as well. The general solution to the Schrödinger equation is

$$\phi(x) = \begin{cases} Aexp(ik_1x) + Bexp(-ik_1x) \\ Cexp(ik_2x) + Dexp(-ik_2x) \end{cases} \tag{8.9}$$

Imposing the continuity conditions at the step for the wave function and its first derivative, we obtain two relations linking the four coefficient of equation 8.9

$$\begin{aligned} C &= \frac{1}{2}\left(1 + \frac{k_1}{k_2}\right)A + \frac{1}{2}\left(1 - \frac{k_1}{k_2}\right)B \\ D &= \frac{1}{2}\left(1 - \frac{k_1}{k_2}\right)A + \frac{1}{2}\left(1 + \frac{k_1}{k_2}\right)B \end{aligned} \tag{8.10}$$

This can be related to the simpler situation of figure 8.4 by setting $A = 1$, $B = r$, $C = t$ and $D = 0$, which gives

$$t = \frac{2k_1}{k_1 + k_2}, r = \frac{k_1 - k_2}{k_1 + k_2}. \tag{8.11}$$

If $E < E_2$, the waves to the right of the step are evanescent (real, decaying exponentials) with a wave number κ_2 given by $\kappa_2^2 = 2m(E_2 - E)/\hbar$.

Equation 8.10 can be written in a matrix form as following:

$$\begin{pmatrix} C \\ D \end{pmatrix} = T \begin{pmatrix} A \\ B \end{pmatrix} \tag{8.12}$$

where the matrix T is given by

$$T = \frac{1}{2k_2}\begin{pmatrix} k_2 + k_1 & k_2 - k_1 \\ k_2 - k_1 & k_2 + k_1 \end{pmatrix} \tag{8.13}$$

Of course this is for $E > E_2$; we need to set $k_2 = i\kappa_2$ if $E < E_2$.

The matrix T is what we call a T-matrix. Its main property is that it gives the coefficients of the wave on the right of the step as a function of those on the left.

The reason for choosing T-matrices is that they can easily be multiplied to build up complicated barriers in one dimension. Infact a potential step can be used as approximated building block for any possible band profile.

The preceding matrix is for a step at the origin and it will be helpful to generalise it to a step at $x = d$. The only difference lies in the phase of the waves that hit the object. Using simple algebraic transformations we can link the generic T-matrix calculated at the distance $x = d$ with the T-matrix at $x = 0$ using the following:

$$T(d) = \begin{pmatrix} exp(-ik_2d) & 0 \\ 0 & exp(ik_2d) \end{pmatrix} T(0) \begin{pmatrix} exp(ik_1d) & 0 \\ 0 & exp(-ik_1d) \end{pmatrix} \tag{8.14}$$

In case of an heterostructure the previous derivation is not strictly correct since the matching conditions at the heterointerface must use the derivative divided by the effective mass, not just the derivative by itself. Furthermore, it is important to normalise propagating states such that they carry equal current. To do this, the coefficient A, B, C and D in equation 8.9 have to be multiply by the factor $\sqrt{\frac{m}{\hbar k_1}}$, where m is the effective mass of the electron in the material where the wave is considered.

The general theory of T-matrices applies to any barrier since a smoothly varying barrier can be handled by approximating its shape as a series of steps. The overall T-matrix is simply given by multiplying those for all the steps. However this technique it is not numerically stable when complicate potential profiles require multiplication of many T-matrices. The problem becomes worse when tunneling barriers are part of the system to solve. Here exponentially growing

and decaying solutions are both possible even if numerical errors may cause the exponential solution to rapidly dominate the wave function.